Lecture Notes in Physics

702

J. Ehlers
C. Lämmerzahl
(Eds.)

Special Relativity

Will it Survive the next 101 Years?

Springer

Lecture Notes in Physics

Editorial Board

R. Beig, Wien, Austria
W. Beiglböck, Heidelberg, Germany
W. Domcke, Garching, Germany
B.-G. Englert, Singapore
U. Frisch, Nice, France
P. Hänggi, Augsburg, Germany
G. Hasinger, Garching, Germany
K. Hepp, Zürich, Switzerland
W. Hillebrandt, Garching, Germany
D. Imboden, Zürich, Switzerland
R. L. Jaffe, Cambridge, MA, USA
R. Lipowsky, Golm, Germany
H. v. Löhneysen, Karlsruhe, Germany
I. Ojima, Kyoto, Japan
D. Sornette, Nice, France, and Zürich, Switzerland
S. Theisen, Golm, Germany
W. Weise, Garching, Germany
J. Wess, München, Germany
J. Zittartz, Köln, Germany

The Lecture Notes in Physics

The series Lecture Notes in Physics (LNP), founded in 1969, reports new developments in physics research and teaching – quickly and informally, but with a high quality and the explicit aim to summarize and communicate current knowledge in an accessible way. Books published in this series are conceived as bridging material between advanced graduate textbooks and the forefront of research to serve the following purposes:

• to be a compact and modern up-to-date source of reference on a well-defined topic;
• to serve as an accessible introduction to the field to postgraduate students and nonspecialist researchers from related areas;
• to be a source of advanced teaching material for specialized seminars, courses and schools.

Both monographs and multi-author volumes will be considered for publication. Edited volumes should, however, consist of a very limited number of contributions only. Proceedings will not be considered for LNP.

Volumes published in LNP are disseminated both in print and in electronic formats, the electronic archive is available at springerlink.com. The series content is indexed, abstracted and referenced by many abstracting and information services, bibliographic networks, subscription agencies, library networks, and consortia.

Proposals should be sent to a member of the Editorial Board, or directly to the managing editor at Springer:

Dr. Christian Caron
Springer Heidelberg
Physics Editorial Department I
Tiergartenstrasse 17
69121 Heidelberg/Germany
christian.caron@springer.com

Jürgen Ehlers Claus Lämmerzahl (Eds.)

Special Relativity

Will it Survive the Next 101 Years?

Springer

Editors

Jürgen Ehlers
Albert-Einstein-Institut
MPI Gravitationsphysik
Am Mühlenberg 1
14476 Golm, Germany
E-mail: mpoessel@aei-potsdam.mpg.de

Claus Lämmerzahl
ZARM, Universität Bremen
Am Fallturm
28359 Bremen, Germany
E-mail: laemmerzahl@zarm.uni-bremen.de

J. Ehlers and C. Lämmerzahl, *Special Relativity*,
Lect. Notes Phys. 702 (Springer, Berlin Heidelberg 2006), DOI 10.1007/b11758914

Library of Congress Control Number: 2006928275

ISSN 0075-8450
ISBN-10 3-540-34522-1 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-34522-0 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable for prosecution under the German Copyright Law.

Springer is a part of Springer Science+Business Media
springer.com
© Springer-Verlag Berlin Heidelberg 2006

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Typesetting: by the authors and techbooks using a Springer LaTeX macro package
Cover design: *design & production* GmbH, Heidelberg

Printed on acid-free paper SPIN: 11758914 54/techbooks 5 4 3 2 1 0

Preface

Einstein's relativity theories changed radically the physicists' conception of space and time. The Special Theory, i.e., Minkowski spacetime and Poincaré-invariance, not only removed an inconsistency between the kinematical foundations of mechanics and electrodynamics but provided a framework for all of physics except gravity. Even General Relativity kept the most essential ingredient of special relativity – a Lorentz-metric – and, therefore, maintained Lorentz-invariance infinitesimally. In the large realm of particle physics where intrinsic, tidal gravitational fields are totally negligible, Poincaré-invariance combined with gauge invariance led to relativistic quantum field theories and, specifically, to the standard model of particle physics.

General Relativity theory and Quantum Field theory generalized classical Poincaré-invariant field theory in different directions. Both generalizations turned out to be successful, but their basic assumptions contradict each other. Attempts to overcome this "most glaring incompatibility of concepts" (F. Dyson) so far have led to partial successes but not to a unified foundation of physics encompassing gravity and quantum theory. Thus, after about a century of successes in separate areas, physicists feel the need to probe the limits of validity of the SR-based theories. Canonical approaches to quantum gravity, non-commutative geometry, (super-)string theory, and unification scenarios predict tiny violations of Lorentz-invariance at high energies. Accordingly, the present seminar tries to cover the basics of Special Relativity, proposed scenarios that lead to violations of Lorentz-invariance, and experiments designed to find such effects. Furthermore, some historical and philosophical aspects are treated.

The main topis of this seminar are

- The foundations and the mathematics of Special Relativity
- Conjectured violations of Lorentz-invariance
- Confrontation with high-precision experiments
- Philosophical and historical aspects

The 271st WE–Heraeus Seminar on Special Relativity, where these issues have been discussed, took place in Potsdam from February 13–18, 2005. We

sincerely thank all speakers for their presentations and especially those who moreover were willing to write them up for the present volume. Last but not least we thank the Wilhelm and Else Heraeus Foundation for its generous support, without which this seminar could not have been realized.

Golm and Bremen *Jürgen Ehlers*
January 2006 *Claus Lämmerzahl*

Experimental set-up of an early high precision search for an anisotropy of inertia.

Contents

Part I Historical and Philosophical Aspects

Isotropy of Inertia: A Sensitive Early Experimental Test
R.W.P. Drever 3
1 Introduction 3
2 Early Ideas 4
3 Possibilities for Experiments 4
4 Some Factors Expected to Affect Sensitivity in a Simple NMR Measurement 5
5 Development of the Experimental Technique 5
6 Initial Observations 7
7 Experiments and Developments for Higher Sensitivity 7
8 Experimental Procedure 9
9 Discussion of Experimental Results 12
10 Interpretation 12
11 Some Personal Remarks 13
References 13

The Challenge of Practice: Einstein, Technological Development and Conceptual Innovation
M. Carrier 15
1 Knowledge and Power in the Scientific Revolution 15
2 Contrasting Intuitions on the Cascade Model 17
3 Poincaré, Einstein, Distant Simultaneity, and the Synchronization of Clocks 20
4 The Emerging Rule of Global Time 24
5 Technology-Based Concepts and the Rise of Operationalism 25
6 Technological Problems, Technological Solutions, and Scientific Progress 28
References 30

Part II Foundation and Formalism

Foundations of Special Relativity Theory
J. Ehlers . . . 35
1 Introduction . . . 35
2 Inertial Frames . . . 36
3 Poincaré Transformations . . . 36
4 Minkowski Spacetime . . . 39
5 Axiomatics . . . 40
6 The Principle of Special Relativity and Its Limits . . . 40
7 Examples . . . 41
8 Accelerated Frames of Reference . . . 41
9 SR Causality . . . 42
References . . . 43

Algebraic and Geometric Structures in Special Relativity
D. Giulini . . . 45
1 Introduction . . . 45
2 Some Remarks on "Symmetry" and "Covariance" . . . 46
3 The Impact of the Relativity Principle on the Automorphism Group of Spacetime . . . 49
4 Algebraic Structures of Minkowski Space . . . 55
5 Geometric Structures in Minkowski Space . . . 71
A Appendices . . . 98
References . . . 108

Quantum Theory in Accelerated Frames of Reference
B. Mashhoon . . . 112
1 Introduction . . . 112
2 Hypothesis of Locality . . . 113
3 Acceleration Tensor . . . 115
4 Nonlocality . . . 116
5 Inertial Properties of a Dirac Particle . . . 119
6 Rotation . . . 120
7 Sagnac Effect . . . 121
8 Spin-Rotation Coupling . . . 122
9 Translational Acceleration . . . 125
10 Discussion . . . 129
References . . . 129

Vacuum Fluctuations, Geometric Modular Action and Relativistic Quantum Information Theory
R. Verch . . . 133
1 Introduction . . . 133
2 From Quantum Mechanics and Special Relativity to Quantum Field Theory . . . 137

3 The Reeh–Schlieder–Theorem and Geometric Modular Action . . . 146
4 Relativistic Quantum Information Theory: Distillability in Quantum Field Theory . . . 154
References . . . 160

Spacetime Metric from Local and Linear Electrodynamics: A New Axiomatic Scheme
F.W. Hehl and Y.N. Obukhov . . . 163
1 Introduction . . . 163
2 Spacetime . . . 164
3 Matter – Electrically Charged and Neutral . . . 165
4 Electric Charge Conservation . . . 166
5 Charge Active: Excitation . . . 166
6 Charge Passive: Field Strength . . . 167
7 Magnetic Flux Conservation . . . 168
8 Premetric Electrodynamics . . . 168
9 The Excitation is Local and Linear in the Field Strength . . . 170
10 Propagation of Electromagnetic Rays ("Light") . . . 173
11 No Birefringence in Vacuum and the Light Cone . . . 175
12 Dilaton, Metric, Axion . . . 180
13 Setting the Scale . . . 181
14 Discussion . . . 182
15 Summary . . . 184
References . . . 184

Part III Violations of Lorentz Invariance?

Overview of the Standard Model Extension: Implications and Phenomenology of Lorentz Violation
R. Bluhm . . . 191
1 Introduction . . . 191
2 Motivations . . . 194
3 Constructing the SME . . . 197
4 Spontaneous Lorentz Violation . . . 203
5 Phenomenology . . . 212
6 Tests in QED . . . 215
7 Conclusions . . . 221
References . . . 222

Anything Beyond Special Relativity?
G. Amelino-Camelia . . . 227
1 Introduction and Summary . . . 227
2 Some Key Aspects of Beyond-Special-Relativity Research . . . 232
3 More on the Quantum-Gravity Intuition . . . 239
4 More on the Quantum-Gravity-Inspired DSR Scenario . . . 244

5 More on the Similarities with Beyond-Standard-Model Research 272
6 Another Century? 274
References 275

Doubly Special Relativity as a Limit of Gravity
K. Imiłkowska and J. Kowalski-Glikman 279
1 Introduction 279
2 Postulates of Doubly Special Relativity 280
3 Constrained BF Action for Gravity 284
4 DSR from 2+1 Dimensional Gravity 290
5 Conclusions 295
References 296

Corrections to Flat-Space Particle Dynamics Arising from Space Granularity
L.F. Urrutia 299
1 Introduction 299
2 Basic Elements from Loop Quantum Gravity (LQG) 304
3 A Kinematical Estimation of the Semiclassical Limit 312
4 Phenomenological Aspects 318
References 340

Part IV Experimental Search

Test Theories for Lorentz Invariance
C. Lämmerzahl 349
1 Introduction 349
2 Test Theories 351
3 Model-Independent Descriptions of LI Tests 354
4 The General Frame for Kinematical Test Theories 364
5 The Test Theory of Robertson 367
6 The General Formalism 376
7 The Mansouri-Sexl Test Theory 379
8 Discussion 381
References 383

Test of Lorentz Invariance Using a Continuously Rotating Optical Resonator
S. Herrmann, A. Senger, E. Kovalchuk, H. Müller, A. Peters 385
1 Introduction 385
2 Setup 387
3 LLI-Violation Signal According to SME 389
4 LLI-Violation Signal According to RMS 394
5 Data Analysis 396
6 Outlook 398
References 400

A Precision Test of the Isotropy of the Speed of Light Using Rotating Cryogenic Optical Cavities
S. Schiller, P. Antonini, M. Okhapkin 401
1 Introduction 401
2 Experimental Setup 402
3 Characterization of the Setup 407
4 Data Collection and Analysis 410
5 Conclusions 413
References 414

Rotating Resonator-Oscillator Experiments to Test Lorentz Invariance in Electrodynamics
M. E. Tobar, P.L. Stanwix, M. Susli, P. Wolf, C.R. Locke, E.N. Ivanov . . 416
1 Introduction 416
2 Common Test Theories to Characterize Lorentz Invariance 417
3 Applying the SME to Resonator Experiments 424
4 Comparison of Sensitivity of Various Resonator Experiments in the SME 433
5 Applying the RMS to Whispering Gallery Mode Resonator Experiments 437
6 The University of Western Australia Rotating Experiment 439
7 Data Analysis and Interpretation of Results 445
8 Summary 448
References 450

Recent Experimental Tests of Special Relativity
P. Wolf, S. Bize, M.E. Tobar, F. Chapelet, A. Clairon, A.N. Luiten, G. Santarelli 451
1 Introduction 451
2 Theoretical Frameworks 452
3 Michelson-Morley and Kennedy-Thorndike Tests 459
4 Atomic Clock Test of Lorentz Invariance in the SME Matter Sector 468
5 Conclusion 475
References 477

Experimental Test of Time Dilation by Laser Spectroscopy on Fast Ion Beams
G. Saathoff, G. Huber, S. Karpuk, C. Novotny, S. Reinhardt, D. Schwalm, A. Wolf, G. Gwinner 479
1 Introduction 479
2 Principle of the Ives Stilwell Experiment 480
3 Ives-Stilwell Experiment at Storage Rings 481
4 Outlook 490
References 492

Tests of Lorentz Symmetry in the Spin-Coupling Sector
R.L. Walsworth 493
1 Introduction 493
2 ^{129}Xe/^{3}He maser
(Harvard-Smithsonian Center for Astrophysics) 494
3 Hydrogen Maser
(Harvard-Smithsonian Center for Astrophysics) 497
4 Spin-Torsion Pendula
(University of Washington and Tsing-Hua University) 499
5 K/^{3}He Co-Magnetometer
(Princeton University) 502
References 504

Do Evanescent Modes Violate Relativistic Causality?
G. Nimtz 506
1 Introduction 506
2 Wave Propagation 508
3 Photonic Barriers, Examples of Evanescent Modes 510
4 Evanescent Modes Are not Observable 515
5 Velocities, Delay Times, and Signals 516
6 Partial Reflection: An Experimental Method
to Demonstrate Superluminal Signal Velocity
of Evanescent Modes 522
7 Evanescent Modes a Near Field Phenomenon 524
8 Superluminal Signals Do not Violate Primitive Causality 527
9 Summary 529
References 530

List of Contributors

Giovanni Amelino–Camelia
Dipartimento di Fisica
Universitá di Roma "La Sapienza" and
Sez. Roma1 INFN
P.le Moro 2
00185 Roma
Italy
amelino@roma1.infn.it

Piergiogio Antonini
Institut für Experimentalphysik
Heinrich–Heine–Universität Düsseldorf
40225 Düsseldorf
Germany
piergiorgio.antonini@
uni-duesseldorf.de

Sébastien Bize
BNM-SYRTE
Observatoire de Paris
61 avenue de l'Observatoire
75014 Paris
France
sebastien.bize@obspm.fr

Robert Bluhm
Colby College
Waterville ME 04901
USA
rtbluhm@colby.edu

Martin Carrier
University of Bielefeld
Faculty of the History of Science
Philosophy and Theology
Department of Philosophy
P.O.B. 100131
33501 Bielefeld
Germany
mcarrier@philosophie.uni-bielefeld.de

Fréderic Chapelet
BNM-SYRTE
Observatoire de Paris
61 avenue de l'Observatoire
75014 Paris
France
frederic.chapelet@obspm.fr

André Clairon
BNM-SYRTE
Observatoire de Paris
61 avenue de l'Observatoire
75014 Paris
France
andre.clairon@obspm.fr

Ronald W.P. Drever
California Institute of Technology, 200-36
Pasadena, CA 91125
USA
rdrever@caltech.edu

Jürgen Ehlers
Max–Planck–Institut für Gravitations-
physik
Albert–Einstein–Institut
Am Mühlenberg
14476 Golm
Germany
juergen.ehlers@aei.mpg.de

Domenico Giulini
Physikalisches Institut
Universität Freiburg
Hermann–Herder–Str. 3
79104 Freiburg
Germany
giulini@physik.uni-freiburg.de

Gerald Gwinner
Department of Physics & Astronomy
University of Manitoba
Winnipeg, MB R3T 2N2
Canada
ginner@physics.umanitoba.ca

Friedrich W. Hehl
Institute for Theoretical Physics
University of Cologne
50923 Köln
Germany
hehl@thp.uni-koeln.de

Sven Herrmann
Institut für Physik
Humboldt Universität zu Berlin
10117 Berlin
Germany
sven.herrmann@physik.hu-berlin.de

Gerhard Huber
Institut für Physik
Universität Mainz
55099 Mainz
Germany
gerhard.huber@uni-mainz.de

Katarzyna Imiłkowska
Institute of Theoretical Physics
University of Wrocław
Pl. Maxa Borna 9
50-204 Wrocław
Poland
kaim@ift.uni.wroc.pl

Eugene N. Ivanov
University of Western Australia
School of Physics M013
35 Stirling Hwy.
Crawley 6009 WA
Australia
eugene@physics.uwa.edu.au

Sergej Karpuk
Institut für Physik
Universität Mainz
55099 Mainz
Germany
karpuk@uni-mainz.de

Evgeny Kovalchuk
Institut für Physik
Humboldt Universität zu Berlin
10117 Berlin
Germany
evgeny.kovalchuk@physik.hu-berlin.de

Jerzy Kowalski–Glikman
Institute of Theoretical Physics
University of Wrocław
Pl. Maxa Borna 9
50-204 Wrocław
Poland
jurekk@ift.uni.wroc.pl

Claus Lämmerzahl
ZARM
Universität Bremen
Am Fallturm
28359 Bremen
Germany
laemmerzahl@zarm.uni-bremen.de

Clayton R. Locke
University of Western Australia
School of Physics M013
35 Stirling Hwy.
Crawley 6009 WA
Australia
clocke@physics.uwa.edu.au

André N. Luiten
University of Western Australia
School of Physics
Nedlands 6907 WA
Australia
andre@pd.uwa.edu.au

Bahram Mashhoon
Department of Physics and Astronomy
University of Missouri–Columbia
Columbia, Missouri 65211
USA
mashhoonb@missouri.edu

Holger Müller
Physics Department
Stanford University
Stanford, CA 94305
USA
holgerm@stanford.edu

Günter Nimtz
II. Physikalisches Institut
Universität zu Köln
Zülpicher Str. 77
50937 Köln
Germany
G.Nimtz@uni-koeln.de

Christian Novotny
Institut für Physik
Universität Mainz
55099 Mainz
Germany
christian.novotny@uni-mainz.de

Yuri N. Obukhov
Department of Theoretical Physics
Moscow State University
117234 Moscow
Russia
yo@thp.uni-koeln.de

Maxim Okhapkin
Institut für Experimentalphysik
Heinrich–Heine–Universität Düsseldorf
40225 Düsseldorf
Germany
Okhapkin@uni-duesseldorf.de

Achim Peters
Institut für Physik
Humboldt Universität zu Berlin
10117 Berlin
Germany
achim.peters@physik.hu-berlin.de

Sascha Reinhardt
Max-Planck-Institut für Kernphysik
69029 Heidelberg
Germany
sascha.reinhardt@mpi-hd.mpg.de

Guideo Saathoff
Max-Planck-Institut für Kernphysik
69029 Heidelberg
Germany
guido.saathoff@mpi-hd.mpg.de

Giorgio Santarelli
BNM-SYRTE
Observatoire de Paris
61 avenue de l'Observatoire
75014 Paris
France
giorgio.santarelli@obspm.fr

Stephan Schiller
Institut für Experimentalphysik
Heinrich–Heine–Universität Düsseldorf
40225 Düsseldorf
Germany
step.schiller@uni-duesseldorf.de

Dirk Schwalm
Max-Planck-Institut für Kernphysik
69029 Heidelberg
Germany
dirk.schwalm@mpi-hd.mpg.de

Alexander Senger
Institut für Physik
Humboldt Universität zu Berlin
10117 Berlin
Germany
alexander.senger@physik.hu-berlin.de

Paul L. Stanwix
University of Western Australia
School of Physics M013
35 Stirling Hwy.
Crawley 6009 WA
Australia
pstanwix@physics.uwa.edu.au

Mohamad Susli
University of Western Australia
School of Physics M013
35 Stirling Hwy.
Crawley 6009 WA
Australia
suslim01@tartarus.uwa.edu.au

Michael E. Tobar
University of Western Australia
School of Physics M013
35 Stirling Hwy.
Crawley 6009 WA
Australia
mike@physics.uwa.edu.au

Luis F. Urrutia
Instituto de Ciencias Nucleares
Universidad Nacional Autónoma de México
Circuito Exterior, C.U.
04510 México, D.F.
México
urrutia@nucleares.unam.mx

Rainer Verch
Institut fr Theoretische Physik
Universität Leipzig
Postfach 10 09 20
04009 Leipzig
Germany
verch@itp.uni-leipzig.de

Ronald L. Walsworth
Harvard–Smithsonian Center for Astrophysics
Cambridge, MA 02138
USA
rwalsworth@cfa.harvard.edu

Andreas Wolf
Max-Planck-Institut für Kernphysik
69029 Heidelberg
Germany
andreas.wolf@mpi-hd.mpg.de

Peter Wolf
BNM-SYRTE
Observatoire de Paris
61 avenue de l'Observatoire
75014 Paris
France
peter.wolf@obspm.fr

Part I

Historical and Philosophical Aspects

Isotropy of Inertia: A Sensitive Early Experimental Test

R.W.P. Drever

California Institute of Technology, 200-36, Pasadena, CA 91125, USA
rdrever@caltech.edu

Abstract. An experimental test for anisotropy of inertia performed by a nuclear free-precession experiment is described. The precession was observed in the Earth's magnetic field, in a countryside location in the open air. The experiment was exceptionally sensitive, and slightly unusual in other ways. Some of the background and other aspects are briefly discussed.

1 Introduction

When I was asked to give an account of an early experiment[1] on "Isotropy of Inertia" which I conceived and carried out many years ago I was reluctant at first. Then I realized that there might be some usefulness, and possibly interest, in this since the experiment was unusual in several ways, and was very different from typical experiments done now. And it might be interesting to explain how some of the ideas arose, and how some of the problems were overcome, in a more personal way than usual.

This experiment was conceived and carried out around 1960, at a time when I was working on experimental nuclear physics in the Natural Philosophy (physics) Department of the University of Glasgow, in Scotland. I had obtained a Ph.D. a few years earlier for work relating to low energy beta spectroscopy and other research on radioactive nuclei carried out using special gas proportional counter techniques developed for the purpose. I was, however, also interested in possibilities of experimental work relating to cosmology, and in the book on cosmology by H. Bondi [1] had come across the suggestion that a test of Mach's Principle ideas on inertia might be possible by looking for some anisotropy in inertial mass. If the inertial mass of a body on the Earth arose from coupling to all other matter in the universe, then the Earth's position to one side of the centre of our galaxy might lead to some anisotropy in the inertial mass of bodies on the Earth. A fairly specific hypothesis of this kind was that of Kaempffer [2], which

[1] Experiment performed (in 1960/61) while at: Department of Natural Philosophy, University of Glasgow, Glasgow, G12 9QD, Scotland

R.W.P. Drever: *Isotropy of Inertia: A Sensitive Early Experimental Test*, Lect. Notes Phys. **702**, 3–14 (2006)
DOI 10.1007/3-540-34523-X_1 © Springer-Verlag Berlin Heidelberg 2006

I found quite appealing. Cocconi and Salpeter [3] took the general idea further by estimating possible shifts of atomic energy levels, and set an upper limit to mass anisotropy from this.

2 Early Ideas

At around this time I realized that similar effects could show up in suitable nuclei, and these could set more sensitive limits since the nuclear building energies involved are so much larger than the binding of electrons in atoms. Cocconi and Salpeter realized this also, and suggested [4] use of the Mössbauer Effect to measure this. It had occurred to me that more sensitive and direct measurements could be made by measuring transitions between levels involving predominantly different distributions of nucleon momentum, using nuclear magnetic resonance techniques (NMR). In fact I found it possible to set new upper limits to anisotropic effects from the width of published NMR resonances already measured with spin 3/2 nuclei for other purposes.

This finding seemed to me to be worth publishing, and I wrote a brief note on it and submitted it to a major letters journal. My manuscript was returned to me with a comment from the Editor saying that the idea was a good one, but it was already being investigated in experiments by a group at Yale led by V.W. Hughes.

I was at first very saddened by this rejection, and also by learning that the same idea was already being experimentally investigated by a group which was probably very experienced and almost certainly had much better equipment and resources than were available to me for such an experiment.

3 Possibilities for Experiments

I was keen, however, to attempt some experiment of this type myself, and the knowledge that a group in a major institution must have decided it was worth doing was a strong additional stimulus for me. I started to consider all the experimental possibilities I could think of, and assess the factors likely to limit sensitivity in each.

The simplest kind of experiment seemed to be an NMR measurement of transitions between the levels of a nucleus with spin 3/2 in a uniform magnetic field, as a function of the direction of the magnetic field relative to the direction to the centre of our galaxy. In the absence of any anisotropy there would be four equally-spaced magnetic sublevels, with spins $+3/2$, $+1/2$, $-1/2$, and $-3/2$; giving a single NMR frequency. Cocconi and Salpeter suggested that in the presence of a mass anisotropy it was possible that the levels could be slightly shifted, the $+3/2$ and $-3/2$ levels in one direction, and the $+1/2$ and $-1/2$ levels in the opposite direction. This would split the NMR line into a triplet, with a splitting which would be a function of the direction of the magnetic field relative to the direction of the galactic centre. If the magnet providing the field

was attached to the Earth it would rotate as the Earth rotates, giving splitting which would be modulated with a periodicity related to 24 hours (sidereal time).

4 Some Factors Expected to Affect Sensitivity in a Simple NMR Measurement

Estimating sensitivity of an NMR experiment of this type involves the following considerations:

(a) The width of the observed resonance could set a limit to sensitivity for small splitting. Factors affecting line width include the relaxation time, which for a suitable liquid solution may be several seconds, and variations in the magnetic field over the volume of the sample.
(b) In this particular experiment, the strength of the magnetic field is not as directly significant as in other NMR measurements, since the frequency splitting is independent of the field, and has to be compared with the frequency corresponding to a fixed nuclear binding energy.

Consideration of factor (b) might suggest that using a weak magnetic field might be an advantage in this case, as it is usually easier to reduce the spatial variation of the magnetic field if the absolute value of the field itself is small. In the present application it seemed appropriate to consider use of the magnetic field of the Earth itself. Free precession techniques had been developed for measuring the Earth's magnetic field, and it seemed these might be adapted for this experiment. In a location far from ferromagnetic materials the field can be very uniform. This seemed to give an opportunity for a sensitive and relatively simple experiment to be performed at very low cost. This was the technique developed and used in this research.

It may be mentioned that the idea of using the Earth's magnetic field here was stimulated in part by the fact that in the Honours Natural Philosophy student laboratories in the University there was an Earth's-field free precession system, to help educate (and challenge) some of the students. The problem of finding a location having a sufficiently uniform magnetic field near steel-framed buildings made this experiment difficult for students, but free precession proton signals of short duration could be observed with a sample suspended from a rope between upper floors of two different buildings.[2]

5 Development of the Experimental Technique

The original technique for measuring the Earth's magnetic field used a 250 cm^3 sample of water surrounded by a coil, with its axis perpendicular to the direction of the Earth's field. A current is passed through the coil for a few seconds to

[2] It is thought that this interesting experiment was originally introduced to the student laboratory by Dr. Jack M. Reid.

polarize the magnetic moment arising from the protons in the water, and when the current is suddenly interrupted, the proton field precesses about the Earth's field, generating a signal in the coil which is detected by switching to a suitable amplifier system [5, 6].

The nucleus with spin 3/2 chosen for the present experiment was Li^7. A solution of lithium nitrate in water was found to have a suitable relaxation time of around 4 seconds. The Li^7 precession signal had a frequency of 803 Hz in the local Earth's field, which with protons gave a frequency near 2068 Hz. The lower frequency and relative weakness of the Li signal compared with that from protons made it necessary use a larger sample, of around 2 litres, and a stronger magnetizing field, with current from a bank of 6 lead-acid car batteries. This in turn required a more extensive uniform magnetic field than available near the University laboratories. The equipment was therefore moved to a country location in the village of Bishopton, 12 miles West of Glasgow, in the back garden of the house in which I was living at the time. In this area the direction of Earth's magnetic field dips steeply towards the North, in such a way that it passed within 10° of the centre of the Galaxy once each sidereal day, a convenient situation for this experiment.

A simplified schematic diagram of the overall arrangement as eventually developed is shown in Fig. 1.

The lithium nitrate solution is contained in a polythene bottle, surrounded by the coil used for magnetizing and sensing, shown at the extreme left side of

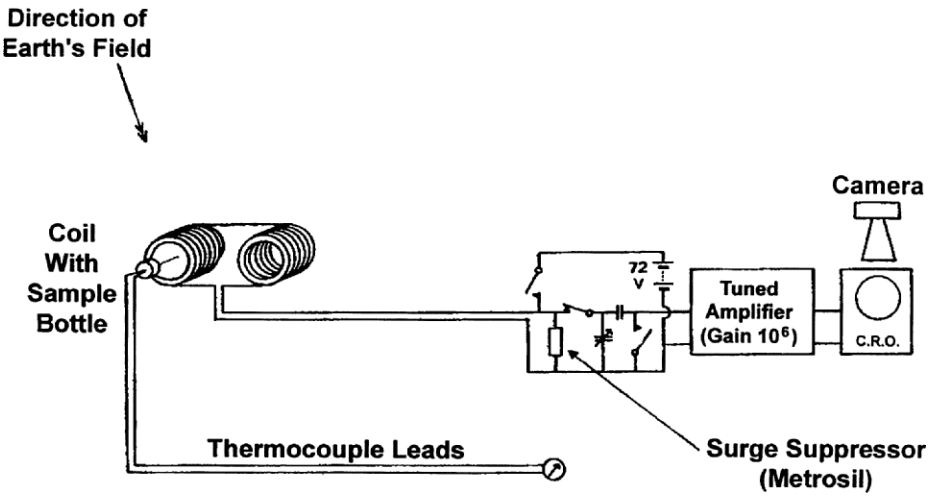


Fig. 1. Simplified diagram of experimental arrangement. Passing a current through the coils produces a net polarization of Li^7 nuclei perpendicular to the direction of the Earth's magnetic field, in a lithium nitrate solution. Rapid switch-off of the current leads to precession of the resulting nuclear magnetization, giving a signal which is examined for beats corresponding to small differential shifts in the nuclear magnetic levels

the figure. The signal was weak, and to minimize interference by electromagnetic fields from the frame time bases of television receivers occasionally operating in the neighborhood, a second similar coil connected in opposition to the sample coil was arranged to cancel signals induced by external magnetic fields. In operation, a magnetizing current is passed through the coils for several seconds to build up a polarization of the nuclear spins perpendicular to the Earth's field. The current is then suddenly turned off, in a time short compared with the precession period, causing the nuclear magnetization to precess about the Earth' field. After a delay of about 0.6 seconds to allow induced voltage transients to decay, the coils are connected to a sensitive tuned amplifier and oscilloscope system to record the free precession signal.

For a single precession frequency, and a uniform magnetic field, the observed signal would be expected to exhibit an exponential decay with a time constant corresponding to the transverse relaxation time of the spin system. If, however the resonance were split into a close triplet it would be expected that the signal would exhibit beats, corresponding to interference between oscillations at the three resonance frequencies which would be detected in a steady-state experiment. A detailed analysis by Das and Saha [7] of the analogous situation of free precession in the presence of a weak electric quadruple interaction indicates that there would be a strong modulation of the signal amplitude at the splitting frequency. If this were due to an anisotropy of inertial mass arising from an interaction with our galaxy it would be expected that the modulation would vary throughout the sidereal day as the direction to the center of our galaxy changes.

6 Initial Observations

The non-uniformity of the Earth's magnetic field in the vicinity of the steel-framed buildings in the Glasgow laboratories had made it very hard to observe free-precession signals from lithium there. However, moving the equipment to the countryside location almost immediately made the lithium precession signals much more detectable. A photographic record of a typical free-precession lithium signal obtained with the arrangement outlined above is shown in Fig. 2. No indication of beating effects of the type expected from anisotropic phenomena were observed at any time, and there were no immediately obvious changes in the records with time of day. Even these initial observations could set better limits to the phenomena being looked for than previous work, and were themselves quite encouraging.

Work then began on a series of further experiments, technical developments, and experimental precautions aimed at improving the sensitivity of the work.

7 Experiments and Developments for Higher Sensitivity

(a) A number of initial tests were made with the coil and sample in various locations, to avoid local non-uniformities of magnetic field. It was found very

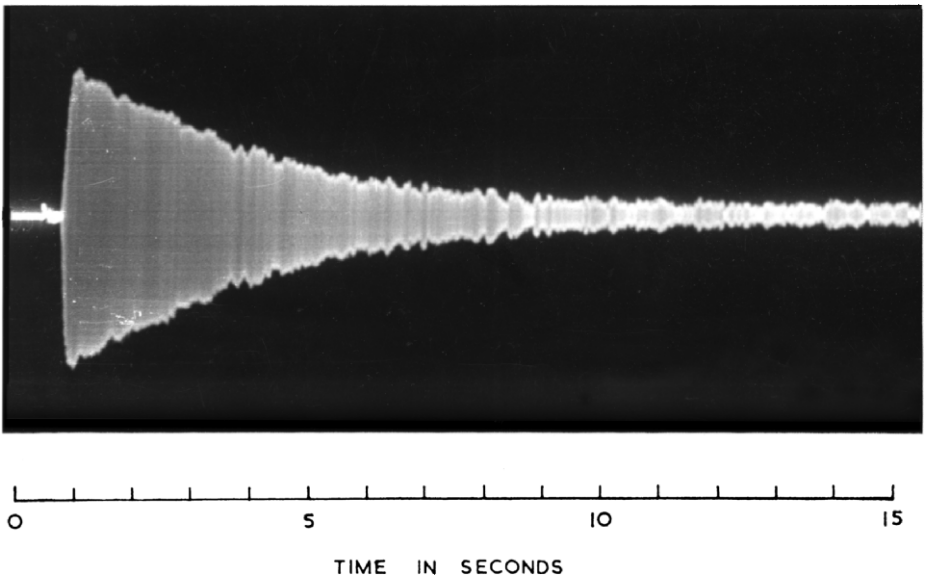


Fig. 2. Typical decay of a free precession signal recorded photographically showing absence of obvious beats over the 15 second time scale indicated

early that allowing the coil to lie directly on the ground gave shorter relaxation times than placing it on a wooden support above the ground. A photograph of an early version of the coil assembly on a wooden metal-free stool in one of the garden locations is shown in Fig. 3, with a close-up view shown in Fig. 4. Tests were also made with the coil assembly supported in the branches of the crab-apple tree seen towards the left side of Fig. 3. No significant difference was observed between the results obtained on the stool and a few metres higher in the tree. Most of the subsequent experiments were made using the wooden stool. The later work was done with the coil assembly nearer the center of a lawn, about 20 m away from the brick wall seen in the background.

(b) The relaxation time in a liquid is a function of temperature, so for observations over 24 hour periods it was important to monitor and control the temperature of the lithium nitrate solution. A later version of the apparatus shown in Fig. 5 incorporates a thermocouple monitor within a polythene sleeve with the end which is inside the bottle sealed. There is also a simple stirring device consisting of a curved copper wire within a similar flexible sealed polythene sleeve. Rotating the wire manually could flex the sleeve, giving effective stirring. In later observations it was arranged that the stirrer could be operated by a small electric motor placed about 20 m from the coil, and coupled to the stirrer by a very light, long belt made from soft medical rubber tubing, 2 mm in diameter. During observations the coil assembly was covered by light plastic sheeting to prevent condensation of dew in the early hours of the morning.

Fig. 3. Photograph of coil and sample bottle on a wooden iron-free stool during early tests in a countryside garden location

(c) To maximize the decay time constant and help keep it constant, nitrogen was bubbled through the lithium nitrate solution to remove dissolved oxygen and the sample bottle was hermetically sealed.
(d) To improve the signal to noise ratio for the lithium precession signal, a slightly more elaborate switching arrangement than that shown in Fig. 1 was eventually used. This involved relays operating in sequence to disconnect and short-circuit the low-noise amplifier system in several places to adequately attenuate the large pulses induced during switch-off of the magnetizing current in the signal coil. A photograph taken during development and testing of the electronic system in one of the teaching laboratories in the University, during a student vacation, is shown in Fig. 6.

8 Experimental Procedure

In operation, a free-precession signal was examined at intervals of 20 or 30 minutes throughout the sidereal day, and photographically recorded using a

Fig. 4. Close-up of the coil and polythene sample bottle

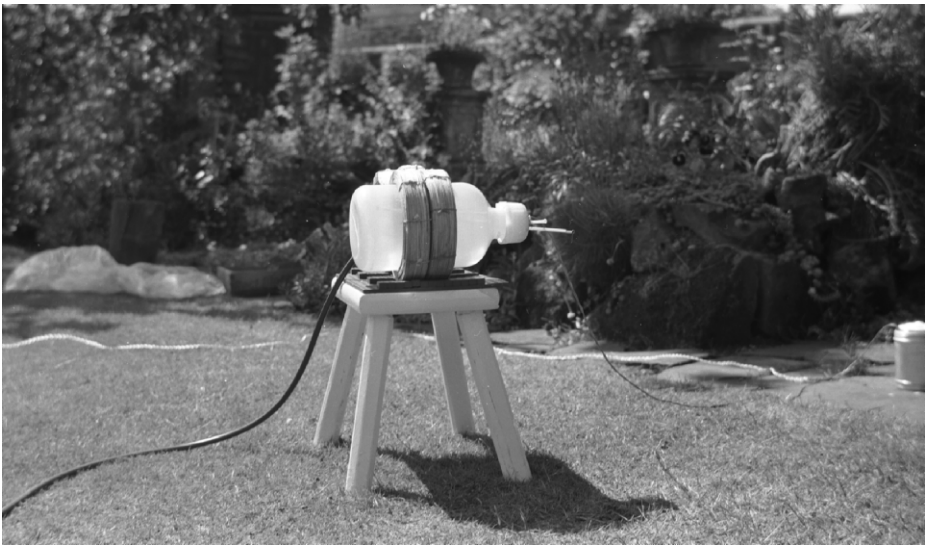


Fig. 5. A later version of the coil and sample system, with a thermocouple temperature monitor. There is a sealed stirrer, operated manually at the time of the photograph and later belt-driven by a small motor from a distance of 20 m. The top of an interference-canceling coil located directly behind the sample coil is just visible

Fig. 6. A photograph taken during development and testing of the electronics and switching system in one of the Honours Natural Philosophy laboratories. A modified low noise nuclear physics amplifier and preamplifier used are on the left and an oscilloscope with a long persistence phosphor on the right. The sample coil was suspended outside the building for these tests

camera with continuously moving film from an oscilloscope with its timebase turned off. The temperature of the lithium nitrate solution was monitored and maintained constant at $37 \pm 1°\mathrm{C}$ by manually adjusting a small current passed through the magnetising coil between the observations.

No sign of a beating pattern or any significant change in the envelope of the precession signal was observed. An upper limit to any effect near the instrumental noise level was determined by projecting the recorded signals onto expected envelope shapes for various assumed energy level shifts. Comparison with a theoretical envelope for the case of a splitting of the resonances by 0.04 Hz, for which the first minimum in the beat pattern occurs near 10 seconds after the start of the precession showed that a slowly varying splitting of this magnitude, which would arise from individual energy level shifts of 0.02 Hz, would have been readily detectable.

This finding alone might not have been enough to completely rule out a much larger effect which moved the outer components of the triplet right outside the pass-band of the amplifier and coil system for most of the sidereal day, allowing them only to pass through the sensitive frequency region at times which happened to coincide with intervals between observations. To check on this unlikely situation a separate experiment was carried out in which the amplitude of the

lithium signal was compared with that from protons in the solution. In doing this it was necessary to take into account the difference in the sensitivity of the apparatus at the two frequencies involved, and to ensure that the decay of the magnetizing field was sufficiently rapid to give maximum signals from both types of nucleus. The experimental results agreed to within 5% of the ratio expected for detection of the whole lithium signal and in disagreement with that expected if only the central component of a triplet had been observed.

It could be concluded that any shifts in the lithium energy levels of the type suggested by Cocconi and Salpeter do not alter the spacing of the levels by more than 0.04 Hz. If one applies the calculation of these authors directly to lithium this would correspond to an upper limit for the ratio of the anisotropic part of inertial mass of the protons involved to its isotropic part of the order of $5 \cdot 10^{-23}$ [8, 9].

9 Discussion of Experimental Results

The high sensitivity achieved in this relatively simple and low-cost experiment was very satisfying for an experimental physicist. And although the earliest published tests for "anisotropy of inertia" were sufficient to rule out effects of magnitude suggested as possible from theories of inertia such as those discussed by Bondi [1], Sciama [10] and Kaempffer [2], the performance of the experiment described here might be taken to correspond to a reserve in sensitivity of as much as a factor of order 10^{15}. This might allow a wide range of second order effects to be ruled out also.

It may be remarked that the use of the Earth's magnetic field did allow this experiment to have significantly higher sensitivity than originally reported from the NMR experiment by the group at Yale University [11], although subsequent improvements in the latter brought its limit [12] closer to that of the experiment described here. It was many years before comparable sensitivity was achieved by other techniques. Major advances in atomic spectroscopy eventually allowed groups at NBS Boulder [13] in 1985, at the University of Washington [14] in 1986, at Harvard University [15] in 1989, and at Amherst College [16] in 1995, to reach even higher precision.

10 Interpretation

The experiment described here was initially partly stimulated by the idea that it might give experimental evidence for or against Mach's Principle, but around the time when early experiments began to give negative results it was pointed out by Epstein [17] that anisotropy in the potential energy of a nucleon could well accompany an anisotropy in its mass and counteract the effects. Dicke [18] showed subsequently that this could be expected, and suggested that these experiments could be regarded as showing, with high precision, that inertial anisotropy effects are universal, the same for all particles. More recently, anisotropy in a variety

of other phenomena and violations of Lorentz invariance in general, have been suggested and experimentally tested. Extensive and more modern discussions of these topics have been given by Will [19], Haugan and Will [20], and others, including contributors to the present Conference Proceedings.

11 Some Personal Remarks

The experiment described here differed in several ways from most current sensitive experiments. As the photographs illustrate, much of the equipment was relatively simple and of low cost, and could be assembled or built fairly quickly.

A strong recollection for me of this work was how exciting it all was to do. A large part of this probably came from the knowledge that the sensitivity was better than anything of the kind known to have been done before, so there was a possibility, even if unlikely, that something quite unexpected and important might be discovered. And moreover a significant result might be found in a 24-hour run, without requiring extensive and time-consuming analysis. The fact that a positive result was not found did not significantly spoil the excitement – something quite new could have shown up.

I might remark also that although I have worked on several very interesting and engrossing kinds of experimental research, this work was by far the most intensely exciting of anything I have been involved in up to now. I say this to encourage others to try to find and work on research that is enjoyable as well as important, which I am sure still exists in all branches of science. If these recollections and comments can be of encouragement to someone I feel that this account will have been well worthwhile.

Acknowledgements

I take this opportunity to acknowledge Philip I. Dee and John C. Gunn, both now deceased, who headed the Department of Natural Philosophy of the University of Glasgow at the time of this work. They gave me their enthusiastic and unstinted encouragement and support for this and my other research over many years, from which I benefited enormously.

References

1. H. Bondi: *Cosmology* (Cambridge University Press, Cambridge 1952), p. 30.
2. F.A. Kaempffer: On possible Realizations of Machs Program, *Canad. J. Phys.* **36**, 151 (1958).
3. G. Cocconi and E. Salpeter: A Search for Anisotropy of Inertia, *Nuovo Cim.* **10**, 646 (1958).
4. G. Cocconi and E. Salpeter: Upper Limit for the Anisotropy of Inertia From the Mossbauer Effect, *Phys. Rev. Lett.* **4**, 176 (1960).

5. M. Packard and R. Varian: Free Nuclear Induction in the Earth's Magnetic Field, *Phys. Rev.* **93**, 941 (1964).
6. G.S. Waters and P.D. Francis: A Nuclear Magnetometer, *J. Sci. Instrum.* **35**, 88 (1958).
7. T.P. Das and A.K. Saha: Electric Quadrupole Interaction and Spin Echoes in Crystals, *Phys. Rev.* **98**, 516 (1955).
8. R.W.P. Drever: Upper Limit to Anisotropy of Inertial Mass from Nuclear Resonance, *Phil. Mag.* **5** (8th S.), 409 (1960).
9. R.W.P. Drever: A Search for Anisotropy of Inertial Mass using a Free Precession Technique, *Phil. Mag.* **6** (8th S.), 683 (1961).
10. D.W. Sciama: On the Origin of Inertia, *Monthly Not. Royal Astr. Soc.* **113**, 34 (1953).
11. V.W. Hughes, H.G. Robinson, and V. Beltran-Lopez: Upper Limit for the Anisotropy of Inertial Mass from Nuclear Resonance Experiments, *Phys. Rev. Lett.* **4**, 342 (1960).
12. V.W. Hughes: Mach's Principle and Experiments on Mass Anisotropy, **in** H.Y. Chiu and W.F. Hoffmann (eds.): *Gravitation and Relativity* (Benjamin, New York 1964), p. 1206.
13. J.D. Prestage, J.J. Bollinger, W.M. Itano, and D.J. Wineland: Limits for Spatial Anisotropy by Use of Nuclear-Spin-Polarized Be^{9+} Ions, *Phys. Rev Lett.* **57**, 2387 (1985).
14. S.K. Lamoreaux, J.P. Jacobs, B.R. Heckel, F.J. Raab, and E.N. Fortson, New Limits on Spatial Anisotropy from Optically Pumped Hg^{201} and Hg^{199}, *Phys. Rev. Lett.* **57**, 3125 (1986).
15. T.E. Chupp, R.J. Hoare, R.A. Loveman, E.R. Oteiza, J.M. Richardson, M.E. Wagshul, and A.K. Thompson: *Phys. Rev. Lett.* **63**, 1541 (1989).
16. C.J. Berglund, L.R. Hunter, D. Krause, Jr., E.O. Prigge, M.S. Ronfeldt, and S.K. Lamoreaux: New Limits on Local Lorentz Invariance from Hg and Cs Magnetometers, *Phys. Rev. Lett.* **75**, 1879 (1995).
17. S.T. Epstein: On the Anisotrophy of Inertia, *Nuovo Cim.* **16**, 587 (1960).
18. R.H. Dicke: Experimental Tests of Mach's Principle, *Phys. Rev. Lett.* **7**, 359 (1961).
19. C.M. Will: *Theory and Experiment in Gravitational Physics* (Cambridge University Press, Cambridge 1993).
20. M.P. Haugan and C.M. Will: Modern Tests of Special Relativity, *Physics Today* **40** (5) 69, (1987).

The Challenge of Practice: Einstein, Technological Development and Conceptual Innovation

M. Carrier

University of Bielefeld, Faculty of the History of Science, Philosophy and Theology, Department of Philosophy, P.O.B. 100131, 33501 Bielefeld, Germany
mcarrier@philosophie.uni-bielefeld.de

1 Knowledge and Power in the Scientific Revolution

The pioneers of the scientific revolution claimed that the developing system of knowledge they envisioned would be distinguished by its practical usefulness. Galileo Galilei, Francis Bacon, and René Descartes agreed that the newly conceived endeavor of unveiling nature's secrets by means of uncovering its lawful regularities would engender practical progress, too. The novel and revolutionary idea was that knowledge of the causes and the laws of nature would pave the way toward technological innovation. As Bacon claimed, inventions bring about supreme benefit to humankind, and this aim is best served by investigating the processes underlying the operations of nature. Knowledge about nature's workings makes it possible to take advantage of its forces [1, I.§129]. In the same vein, Descartes conceived of technology as an application of this novel type of knowledge. The speculative and superficial claims that had made up the erudition of the past had remained barren and had failed to bear practical fruit. The principles of Descartes' own approach, by contrast, promised to afford

> knowledge highly useful in life; and instead of the speculative philosophy taught in the schools, to discover a practical one, by means of which, knowing the force and action of fire, water, air, the stars, the heavens, and all the other bodies that surround us, as distinctly as we know the various crafts of our artisans, we might apply them in the same way to all the uses to which they are apt, and thus render ourselves the lords and possessors of nature. [2, IV.2, p.101]

The scientific revolution was fueled by the prospect of technological progress. Knowledge of the laws of nature was claimed to be the chief road toward the betterment of the human condition. Bacon quite explicitly stated that studying the processes of nature, or, in present-day terms, carrying out fundamental research, is much better suited for ensuring technological invention than mere trial and

M. Carrier: *The Challenge of Practice: Einstein, Technological Development and Conceptual Innovation*, Lect. Notes Phys. **702**, 15–31 (2006)
DOI 10.1007/3-540-34523-X_2 © Springer-Verlag Berlin Heidelberg 2006

error. Fumbling around with some gadgets is of no avail; rather, systematic observation, methodical experimentation, and painstaking analysis constitute the pivot of technology development [1, I.§110, §117, §129].

However, the emerging science of the 17th century completely failed to live up to these ambitions. The declarations of practical relevance were in no way borne out by the rise of applied research. Quite the contrary. The traditional rift between science and technology remained unbridged for centuries. Christopher Wren was both an outstanding architect and a physicist. In particular, he was familiar with the recently discovered Newtonian mechanics which he thought disclosed the blueprint of the universe. However, when he constructed St. Paul's Cathedral in London, Wren exclusively relied on medieval craft rules. The Newtonian laws accounted for the course of celestial bodies and resolved the mystery of the tides, but they offered no help for mastering the challenges of architecture. Likewise, the steam engine was developed in an endless series of trial and error without assistance from scientific theory [3, p. 162–163]. The operation of the engine was understood only decades after the construction had been completed. The grasp of theory only rarely extended to machines and devices.

Around the middle of the 19th century things began to change. Applied science came into being and successfully connected theory and technology. Tinkering and handicraft were gradually replaced by scientific training. Industrial research emerged and scientists and engineers became the key figures in promoting technological progress. Around 1900, Bacon's vision of a science-based technology had finally become reality.

Bacon's conception of the relation between scientific knowledge and technological power is sometimes called the *cascade model*. The idea is that scientific knowledge flows downward to the material world, as it were, and becomes manifest in useful devices. Practical tasks are best solved by bringing fundamental insights to bear. Deliberate intervention in the course of nature demands uncovering nature's machinery, it requires studying the system of rods, gears, and cogwheels nature employs for the production of the phenomena [1, I.§3, I.§110, I.§117, I.§129].

I wish to explore the relationship between pure and applied research. I will begin by outlining consequences of the cascade model and will sketch an alternative, *emergentist* conception. Both approaches agree in suggesting that the concentration on practical problems which is characteristic of large parts of present-day research is detrimental to the epistemic aspirations of science. These concerns are not without justification. Yet examining Albert Einstein's road toward special relativity theory brings an additional message in its train: Taking practical issues into account may stimulate epistemic progress. I will explain that the operational notion of simultaneity that constituted a key element in the conception of special relativity was suggested by the technological background of the period. Technology became heuristically fruitful for scientific theory. My conclusion is that pure science has less to fear from application pressure than is thought in some quarters.

2 Contrasting Intuitions on the Cascade Model

The growth of scientific knowledge leads to the increasing capacity to cope with intricate circumstances and heavily intertwined causal factors, and this improvement also enhances the practical relevance of scientific theory. As a result, the cascade model appears to provide an adequate portrait of the relationship between scientific progress and technology development. In fact, the cascade model was underlined in the so-called Bush-report issued in 1945 [4]. Vannevar Bush had been asked by President Roosevelt to devise an institutional scheme that would make science in the future post-war period most beneficial to the people. The President was interested in how to improve the usefulness of science; he explicitly mentioned the fighting of diseases and the stimulation of economic growth. In his report, Bush placed fundamental research at center stage. As he argued, new products and new jobs can only be created through continuing basic research. Bush gave two reasons. First, the solution of a practical problem may come about as an unexpected consequence of a seemingly remote theoretical principle. Second, innovative approaches to practical problems often originate from an unfamiliar combination of such principles. Both arguments imply that the theoretical resources needed for meeting a technological challenge often cannot be anticipated and specified in advance. As Bush claimed, practical success will frequently result from fundamental insights in fields and subjects apparently unrelated to the problem at hand. The lesson is clear. The royal road to practically successful science is the broad development of basic research. If useful knowledge is to be gained, it is counterproductive to focus on the concrete issues in question. Rather, forgetting about practical ends and doing fundamental research in the entire scientific field is the first step toward practical accomplishments. In the second step, technologically relevant consequences are drawn from these principles; that is, theoretical models for new technical devices and procedures are derived [4].

The message of the Bush report strongly influenced the public understanding of the relationship between basic and applied research. Indeed, there was and still is an element of truth in it. A large number of the technological innovations in the past decades were achieved by bringing theoretical understanding to bear on practical challenges. For instance, the breathtaking decrease in the size of electronic circuits was accomplished by procedures which draw heavily on theories of optics and solid state physics. Similarly, inventions like optical switches or blue light emitting diodes are produced by joining and combining hitherto unconnected laws of physics. Conversely, what amounts to the same, premature applications may come to grief. A case in point is the striking failure of the American systematic program on fighting cancer. This program was launched in 1971 after the model of the Manhattan Project and the Apollo Program; it included a detailed sequence of research steps to be taken in order to advance cancer prevention and therapy. The practical achievements reached were almost insignificant, and this failure is usually attributed to the fact that the fundamental knowledge necessary for developing successful medical treatment was still lacking [5, p. 211–212].

The cascade model has proved its relevance for relativity theory, too. Einstein's fundamental insights into the factors influencing temporal durations figure prominently in the satellite-based global positioning system (GPS). Numerous satellites in the orbit of the earth broadcast signals from which a terrestrial receiver can infer the time at which the signals were sent. By taking into account the velocity of light, the distance to the relevant satellites can be obtained. It is clear that such a procedure is critically dependent on highly accurate clocks in the satellites. At this juncture, distortions highlighted by special and general relativity come into play. Time dilation slows the orbiting clocks down, the weaker gravitational field makes them run faster. Consequently, the clocks need to be manufactured in such a way that they run inaccurately on Earth – and even substantially inaccurate at that. As a matter of fact, in 1977 when the first cesium clock was launched into the orbit, some engineers doubted the appropriateness of such comparatively huge alterations and insisted that the clocks run at their uncorrected terrestrial rate. A relativistic correction mechanism was built in but remained switched off initially. The signals received exhibited precisely the distortion predicted by the joint relativity theories. After 20 days of increasing error, the correction unit was activated – and has remained so ever since [6, p. 285–289].

Thus, relativity theory is attuned to Bush's leitmotif that theoretical principles may gain unexpected practical significance or, conversely speaking, that the solution to practical problems may come from remote theoretical quarters. You never know for sure in advance which particular corner the light of knowledge will illuminate. Yet, on the whole, the picture is not that clear. Other indications point in the opposite direction. Let me contrast the cascade model with contrary considerations.

Underlying the cascade model is a thorough theoretical optimism. Insights into nature's mode of operation extend to include the subtleties of the functioning of engines and gadgets. Theoretical principles are able to capture the fine details of the phenomena on which the appropriateness and reliability of some artifact turns. Within the sciences, such a sanguine attitude is called reductionism. No feature of nature is small enough or remote enough to escape the grip of the fundamental laws. However, scientists do not embrace reductionism univocally. Rather, its prospects remain contentious. In the U.S. debate around 1990 about the usefulness of building a superconducting collider on Texan soil, one of the warring factions, the particle physicists prominently among them, maintained that unveiling the fundamental processes would shed light on phenomena at higher levels of the organization of matter. That is, discoveries in particle physics should help to clarify properties and interactions at the nuclear, atomic or molecular scale. By contrast, the opposing anti-reductionist or emergentist camp featured the specific character of the phenomena at each level of organization. Emergentists deny that insights about quarks or strings will radiate downward, as it were, and have much impact on the clarification of phenomena from atomic or solid state physics.

Actually, these two factions go back to a venerable opposition in the philosophy of nature, the opposition, namely, between Platonism and Aristotelianism. Platonism is committed to the rule of fundamental law; the universal is supposed to pervade the whole of nature. Aristotelianism insists on the basic and unique character of specific cases; the differences among the particulars outweigh their shared features. This latter view has been prominently supported in the last quarter century by Nancy Cartwright. As she argues, the universal claims of overarching laws are specious; such laws fail to gain access to the phenomena with their rich details and variegated traits. Cartwright takes up an example of Otto Neurath who had drawn attention to the embarrassing silence that seizes Newtonian mechanics in the face of the question where a thousand-shilling bill swept away by the wind in Vienna's St. Stephen's square will hit the ground eventually [7, p. 318]. The only way to get a grip on the phenomena is by making use of local models that are tightly locked onto particular problems. Descriptive adequacy is only accomplished by small-scale accounts; comprehensive theories inevitably lose touch with the wealth of the phenomena. The patchwork quilt, not the pyramid, is symbolic of the structure of scientific knowledge [7, p. 322–323].

Such Aristotelian or emergentist approaches are tied up with a new account of the relation between basic and applied science or epistemic and practical research. The cascade model is abandoned; basic research is said to be largely unsuccessful in meeting applied challenges. Rather, practical problems are to be attacked directly; a detour through the basics is unnecessary and superfluous. Fundamental truths only rarely produce technological spin-offs. Applied research needs to rely on its own forces. The heuristic message of emergentism is that the resources available for addressing practical challenges should be allotted to doing research on precisely these practical challenges.

In fact, a closer inspection of the present state of applied research confirms this latter approach. Industrial companies tend to reduce basic research in favor of target-oriented projects which aim at concrete, marketable goods. Take "giant magnetoresistance" as an example. The underlying physical effect was discovered in 1988; it involves huge ("giant") changes of the electrical resistance of systems composed of thin ferromagnetic layers separated by non-ferromagnetic conducting spacer layers. The resistance of such systems is strongly dependent on the direction of magnetization of the ferromagnetic layers which can be altered by applying an external magnetic field. As a result, the electrical resistance of such an array is influenced by an external field, and this dependence can be used to build extremely sensitive magnetic field sensors. Giant magnetoresistance underlies the functioning of today's magnetic read heads; it is used for hard disks or magnetic tapes. It was realized immediately that the effect is based on spin-dependent scattering of electrons, but such a qualitative explanation was insufficient for constructing suitable devices. For technological use, quantitative relations between relevant parameters such as layer thickness or ferromagnetic coupling between layers were needed. Such relations were not provided by theory, but had to be gained experimentally. When it came to building working devices,

the empirical identification of design rules, not the appeal to fundamental laws, were the order of the day [8].

However, if focusing on narrow, practical issues determines the agenda of applied research, and if fitting parameters is among its chief tools, what kind of science will we end up with? Given the dominance of application-oriented research, its methods and procedures can be expected to radiate into the whole of science. Actually, worries about the detrimental impact of applied research on the methodological dignity of science have been articulated frequently. For instance, theoretical physicist John Ziman complained recently that science guided by material interests and commercial goals will lack objectivity and universality ([9, p. 399]; see [8, Sect. 1]). In the same vein, particle physicist Silvan Schweber claimed that "the demand for relevance ... can easily become a source of corruption of the scientific process" [10, p. 40]. According to such voices, science is likely to suffer in methodological respect from the emphasis on practical use. Application dominance jeopardizes the demanding epistemic standards that used to distinguish science; conversely, retaining such standards requires a commitment to truth rather than utility.

These considerations leave us with a stark alternative concerning the structure of applied research. If the cascade model is correct, concentration on practical issues will dry up practical success in the long run. It would mean eating up the seed corn needed for producing future harvest. If the emergentist approach is correct, practical success is best accomplished by focusing on specific issues, but proceeding in this fashion could spoil the epistemic merits of science. Which side is right? Well, it helps to cast a glance at Einstein who worked at the Bern patent office while pondering the electrodynamics of moving bodies.

3 Poincaré, Einstein, Distant Simultaneity, and the Synchronization of Clocks

It is well known that Einstein in his classical 1905 paper on special relativity suggested two principles as the foundation of the theory he was about to develop. First, the principle of relativity according to which all frames of reference in uniform-rectilinear motion are equivalent, not alone with respect to the laws of mechanics but also regarding electrodynamics including optics [11, pp. 26,29]. Second, the statement that the velocity of light is independent of the motion of the light source. This claim was not peculiar to Einstein but rather a theorem of classical electrodynamics, or the "Maxwell–Lorentz theory."

This latter theory implied, however, that the velocity of light should depend on the motion of the observer. In a series of experiments, conducted in part with Edward Morley, Albert Michelson had established that no such dependence was measurable. Surprisingly enough, the velocity of light came out the same for differently moved observers. Yet the assumed variation in the velocity of light was the chief means for determining the state of motion of an observer. Thus it appeared that different frames of uniform-rectilinear motion or inertial motion

could not be distinguished empirically. This failure posed a serious challenge to electrodynamics to which Hendrik Lorentz responded by developing a more sophisticated version of the theory.

The appropriate application of the principles of electrodynamics (such as Maxwell's equations) demanded that the relevant values of "true motion" or motion with respect to the ether be known. True motion should become manifest in a change in the measured speed of light depending on the velocity of the observer. However, the Michelson–Morley experiment showed that no influence of the motion of the observer on electromagnetic quantities could be recognized. Lorentz pursued a two-pronged strategy for coping with this anomaly. First, he introduced a quantity he called "local time" which differs from place to place and is thus distinguished from true, universal time t. Local time t' is obtained from true time t, the velocity v and the position x of the observer, and the velocity of light: $t' = t - vx/c^2$. Lorentz's proposal was to employ local time for ascertaining the electromagnetic properties of moved bodies. Namely, these properties are determined by calculating them for bodies at rest in the ether at the corresponding local time. In other words, the effect of the motion was taken into account by evaluating the relevant quantities at a time different from the true one. Lorentz considered position-dependent local time as a mathematical artifact for transforming electromagnetic quantities and did not expect that local time showed up on anybody's watch. Second, Lorentz introduced a contraction hypothesis according to which bodies were assumed to shrink as a result of their motion through the ether. This length reduction was thought to be produced by the interaction between moved matter and the ether. The resting ether compresses the body in passage through it, and this contraction precisely cancels the effect of the motion on the velocity of light. The change in the velocity of light induced by the motion is precisely compensated – as the Michelson–Morley null result demands. No effect of the motion on the moved body will be registered ([12, pp. 268–270]; [13, p. 482]; [14, p. 10]; [15, pp. 47–48]; [16, pp. 104–113]; see [17, pp. 130–133], [18, p. 78]).

Lorentz provided his contraction hypothesis with a theoretical backing. He assumed that the forces of cohesion that produce the shape and dimensions of a body are electromagnetic in kind (or at least transform like electromagnetic forces) and was able to derive the contraction hypothesis on this basis. The stated conclusion was that "many" phenomena appear in the same way irrespective of the observer's state of motion, which means that Lorentz did not rule out the existence of tangible effects of the motion of bodies through the ether. That is, his improved theoretical framework did not embody a principle of relativity[1].

From 1900 onward, Henri Poincaré modified Lorentz's approach in two important respects. First, Poincaré had suggested in 1898 that temporal notions like duration or simultaneity are not given by the senses but need to be defined. Defining simultaneity is, as he went on to argue, a matter of coordinating distant

[1] [19, p. 8]; [20, p. 48]. In 1912, Lorentz acknowledged in retrospect that his failure to adopt the principle of relativity as a comprehensive and strict law was responsible for the erroneous parts of his earlier treatment [19, p. 10].

clocks. The options he mentioned for this purpose included the use of globally visible astronomical events, clock transport and electric signals sent by the telegraph ([21, pp. 11–12], [6, pp. 32–37, 238–239]). As Poincaré later made more explicit, the method of choice is sending signals crosswise between two distant clocks and adjusting the clock readings accordingly [14, p. 7]. Poincaré's first conceptual breakthrough was to recognize that if signal exchange was employed for synchronizing distant clocks in motion through the ether, an event happening at true time t at one clock will occur at local time t' at the other [22, p. 483]. That is, in contrast to Lorentz's view, local time was not a mere convenience. Rather, Poincaré's idea of establishing distant simultaneity by synchronizing clocks through signal exchange entailed that local time is observable; it is the time reading the moved clock yields. Second, likewise in contrast to Lorentz, Poincaré assumed that there is no way to distinguish bodies in absolute motion; only relative motions are accessible empirically. This means that Poincaré's version of the Maxwell–Lorentz theory incorporated the principle of relativity ([23, pp. 176–177, 186], [6, pp. 45, 277–279]).

Both assumptions are also characteristic of special relativity theory. Einstein supposed as well that local time is the time provided by a moved clock and is thus given in experience, and he also stated that only relative motions are accessible empirically. Yet this superficial agreement hides a deep-seated divergence as to the nature of local time and the conceptual status of the relativity principle. For Poincaré, local time involved a distortion of true time that was due to the motion through the ether. In reality, the velocity of light is different depending on the motion of the observer; the true value is only assumed in the system at rest in the ether. As a result, the correct simultaneity relations are only obtained within this rest system. However, there is no way to know which system is really at rest. Signal synchrony yields mistaken simultaneity relations for systems in true motion but since all clocks are distorted alike and length relations altered correspondingly, the true simultaneity relations cannot be revealed by experience. The simultaneity relations and the yardstick used for their evaluation change in the same way so that the true relations remain hidden. Consequently, for Poincaré, the principle of relativity constituted a theorem of electrodynamics. It was deduced from electrodynamic assumptions, procedures for establishing simultaneity relations, and the forces acting on charged bodies. In addition, the principle was purely epistemic. In nature, there are privileged frames of reference and absolute motions; yet they are concealed from the unbefitting curiosity of human observers ([23, pp. 188–189]; [14, p. 10]).

Einstein dissented on both counts. First, he placed the relativity principle at the top. After a quick reference to the failed attempts to identify states of absolute rest, he immediately jumped to the principle: "We will raise this conjecture (whose intent will from now on be referred to as the 'Principle of Relativity') to the status of a postulate" [11, p. 26]. In contradistinction to Lorentz and Poincaré, the principle was not supposed to be derived but stated as a premise. Second, Einstein did not confine the principle to observable phenomena but extended it to the theoretical description. This is apparent from the famous opening

paragraph of the 1905 paper in which Einstein criticizes an explanatory asymmetry inherent in the then-current electrodynamics: the interaction between a magnet and a coil is treated differently depending on which object is assumed to be in motion. If a coil is moved in a static magnetic field, an electric current is produced through the Lorentz force; if the magnet is moved, the current is generated by induction. The value of the current agrees in both cases, but its emergence is attributed to different causes. Einstein took this conceptual asymmetry to be utterly implausible. In his view, there was but one phenomenon, namely, coil and magnet in relative motion; and one phenomenon demanded one explanation. Consequently, Einstein was not content with the recognition that the attribution of specific states of motion made no observable difference; he required in addition that the theoretical explanation invoked nothing but relative motion.

However, this creative shift was not enough to save the situation but rather gave rise to a great puzzle. The principle of relativity implies that observers in different states of motion measure the same value of the velocity of light. Yet how is it possible, one must ask, that this quantity comes out the same without appeal to any compensating mechanism? Einstein masters this challenge with another creative shift, namely, the adoption of a procedural definition of simultaneity. From Poincaré, Einstein had learned that judgments about simultaneity are to be based on procedures for synchronizing distant clocks. Einstein elaborated this operational approach to simultaneity and proposed to employ light flashes as a means for synchronizing distant clocks. Two distant clocks are said to be synchronous if the transit time of the signal from the one to the other, as given by reading both clocks, equals the transit time in the backward direction. This is tantamount to saying that the two clocks are synchronous if the reflection of the signal at the distant clock, as measured by that clock, is one half of the period which passes between emission and return of the signal, as measured by the clock at the origin ([11, p. 28]; [24, pp. 196–197]).

Einstein went on to demonstrate that the Lorentz–contraction can be explained on this basis. Observers in relative motion who apply this rule will deviate in their judgments about which events are simultaneous. Measuring the length of a moved body involves locating its edges at the same time. Divergent assessments of the prevailing temporal relations will obviously affect the outcome of length measurements. Lorentz–contraction ceases to be a dynamic effect, based on the action of the forces of cohesion, it becomes a metrogenic effect, based on different judgments about simultaneity. Some argumentative steps later Einstein also succeeded in resolving the conceptual asymmetry in electrodynamics that had prompted his initial worries. Special relativity was born.

Einstein's operational approach to simultaneity was the key to success. However, adopting such an approach is by no means a matter of course. On the contrary, placing all one's bets on signal synchrony seems highly dubious in the face of the counterintuitive results this method yields. Imagine the situation: A criterion for assessing simultaneity relations picks different events as simultaneous according to the state of motion of those who bring the criterion to bear.

Simultaneity ceases to be objective and becomes a frame-dependent notion. How to digest such a finding? One might be tempted to argue that the relativity of simultaneity militates against the procedural approach to simultaneity and suggest that the latter be abandoned. Yet Einstein sticked to it – in spite of its seemingly absurd consequences. And the scientific community quickly accepted this move. But why? What is the reason for Einstein's confidence in the operational notion of simultaneity? And why was the scientific community prepared to follow him on this path?

4 The Emerging Rule of Global Time

The procedural approach to simultaneity was first proposed by Poincaré who recommended the telegraph as a preferred means for synchronizing distant clocks. Yet Poincaré advanced his suggestion not as something new and innovative but as "the definition implicitly admitted by the scientists" [21, p. 11]. Peter Galison recently elucidated the vast technological background to this judgment. Standardizing time readings by coordinating distant clocks constituted one of the chief items on the agenda of technology development in the three decades preceding Einstein's wrestling with the issue. One of the reasons was the rapid expansion of the railroad system. Traditionally, the clocks were set on a local or regional basis by using astronomical procedures. That is, clocks were adjusted to the corresponding mean solar time. The spread of a train service operating on a fixed schedule demanded the coordination or unification of the scattered local time zones.

In addition, an early wave of globalization swept through the late 19th century world. Soaring trade and commerce figures and the foundation of colonies worldwide created a demand for unambiguous time regulations and accurate maps. The problem with drawing global maps lay with measuring longitude differences reliably. In general terms, it was clear how to proceed. The time readings of clocks placed at the relevant positions had to be compared and the local deviations be translated into shifts in the east-west direction. However, a comparison of this sort requires that the clocks run in a coordinated fashion. Accordingly, establishing distant synchrony was not a remote subtlety but rather pervaded the web of commerce, technology, and politics of the period.

In fact, the procedure standardly adopted for synchronizing clocks was sending signals. Around 1880, a pneumatic system was in use in Paris. Air pressure pulses raced through pipes underneath the streets and transmitted time signals to public clocks distributed over the city. The delay due to the transit time of the pressure waves ran up to 15 seconds and was corrected by an array of mechanical counteracting devices [6, pp. 93–95].

From the 1880s onward, this clumsy network of pipes war replaced by a system of cables and wires. The signals employed for synchronizing clocks became electrical; the telegraph made its appearance. Electrocoordinated time connected Europe with North America and with the colonies overseas. The subsequent technological step was taken in the early 20th century. It involved employing

radio waves and allowed surveyors to dispense with a costly network of cables across land and sea. Time coordination and longitude determination became feasible worldwide. Distant synchrony was achieved by emitting a radio signal at a known time and adjusting a distant clock accordingly, taking due account of the transit time. Longitude differences were determined on that basis by using two clocks and sending one radio signal from east to west and another one from west to east [6, pp. 184–186].

In the period under consideration, Poincaré served as chief of the French Bureau de Longitude and was familiar with the practical challenges of coordinating clocks; he referred to the crosswise exchange of signals, i.e., the method in practical use in the administration he headed [14, p. 7]. Likewise, this array of two clocks connected by two signals sent back and forth strikingly resembles the arrangement Einstein invoked for the operational introduction of simultaneity. The only difference is that he referred to light rays whereas electrical signals and radio waves were in general use in his period [11, p. 28]. Likewise, Einstein's passing reference to train schedules as a means for illustrating the importance of simultaneity [11, p. 27] gains a significance that is easily missed otherwise. The technical background makes its presence felt strongly.

It is worth remembering, therefore, that Einstein lived in Bern which, by 1905, ran an extensive network of coordinated clocks, see Fig. 1. It is worth noting, too, that Einstein worked as a technical expert in the Swiss patent office. He reviewed and examined patent applications, and clock making was one of the key technologies of the period. A number of applications concerning electrically coordinated clocks passed through the patent office between 1902 and 1905, some of which must have crossed Einstein's desk [6, p. 248]. It is true, Einstein was critical of Newtonian absolute time and similar metaphysical conceptions as a result of his philosophical studies. Reading the works of Hume, Mill, Mach, and Poincaré had prepared him to accept procedural notions of temporal quantities. Yet the adoption of signal synchrony as the basis of distant simultaneity is no doubt strongly influenced by the technology of his time and his daily work in the patent office. Next to Einstein, the philosopher-scientist, stands Einstein, the patent officer-scientist [6, p. 255]. It is at this juncture where we find the sought-for basis of Einstein's seemingly premature confidence in the operational definition of simultaneity. Here lies the justification for retaining signal synchrony despite its prima-facie implausible ramifications and to transform our spatiotemporal notions on that basis.

5 Technology-Based Concepts and the Rise of Operationalism

The upshot is that the technological development of the period contributed to shaping concepts used in highbrow theory. The procedural approach to simultaneity paved the way toward the understanding of the electrodynamics of moving bodies. The underlying operational attitude is found in both Poincaré and

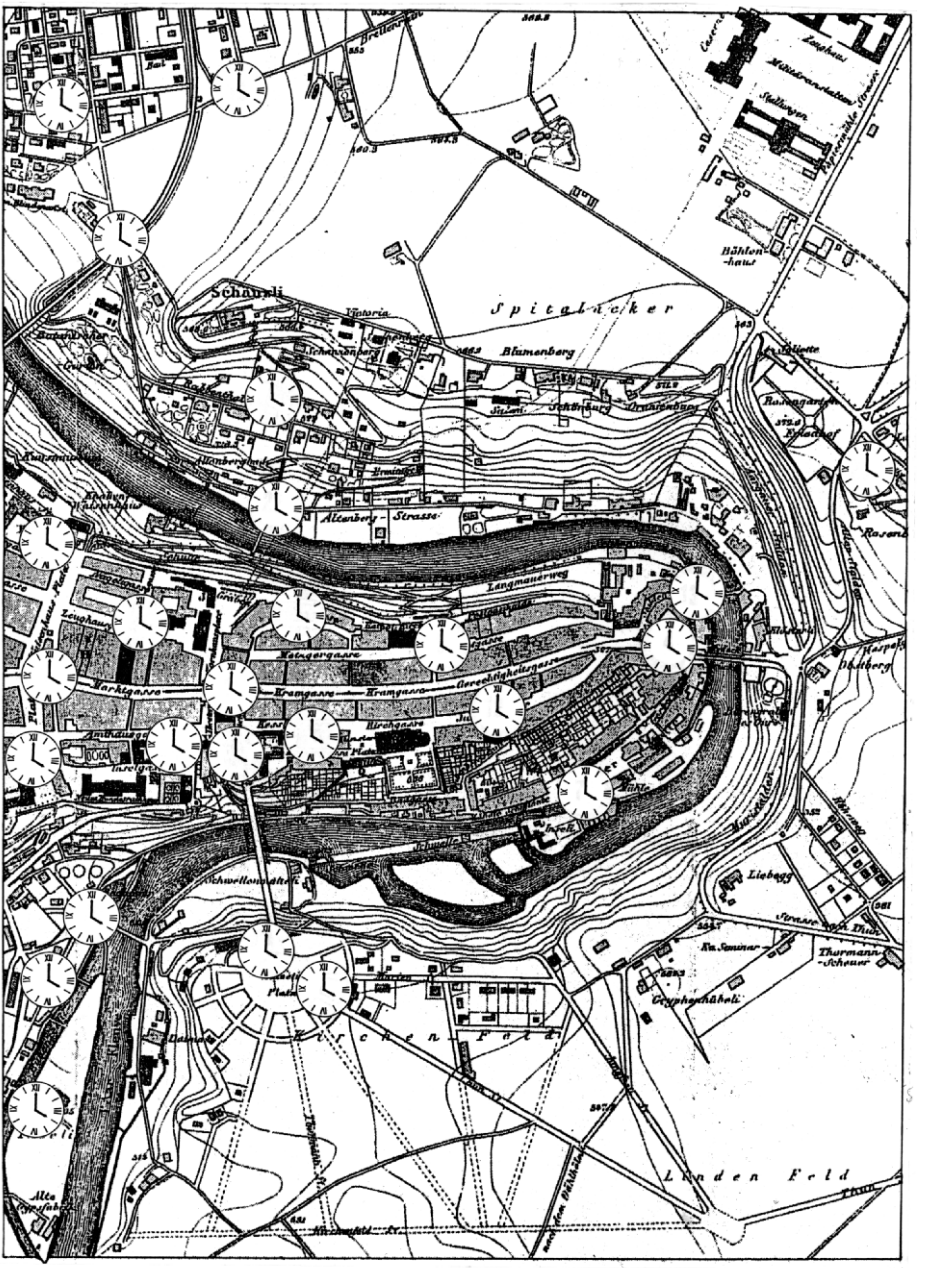


Fig. 1. Bern's Electrical Clock Network by 1905 [26, p. 131] (by courtesy of Chronos–Verlag Zürich)

Einstein, but Einstein pushed this approach much further than Poincaré and thereby prepared the breakthrough to Special Relativity. Poincaré continued to adhere to a privileged, true simultaneity relation. It is true, he emphasized the epistemic problems involved in the identification of true simultaneity. At bottom, Poincaré developed an epistemic circularity argument to the effect that the quantity to be evaluated and the standard used for the evaluation change in the same way so that no observable effect remains. In reality, the velocity of light is influenced by the absolute motion of the observer; but, first, as a result of using signal synchrony and, correspondingly, judging simultaneity relations in terms of local time, and, second, due to the universal contraction of bodies moved through the ether, this influence is invisible in the data.

This means that Poincaré did retain the notion of a preferred frame of reference. The ether rest frame was distinguished among the class of inertial frames in that it alone yields the true measures of lengths, velocities, and electromagnetic quantities. The motion through the ether produces a distortion of these magnitudes which is compensated by other effects of the motion. Poincaré's account, like Lorentz's, involves a sort of conspiracy among different effects brought forth by the motion of bodies. These effects are so contrived as to cancel each other out, hiding in this way the true quantities.

It is characteristic of Poincaré that epistemic problems of this sort did not, in his view, undermine the legitimacy of the concepts involved. The notion of simultaneity remains unaffected. Events happening at the same true time are truly simultaneous – whatever their relation at the local time scale is. Local time is a specious measure of temporal relations; it is flawed by the inability to take absolute motion into account. Likewise, the principle of relativity was confined to the phenomenal realm. In reality, the relevant quantities are affected by the motion, but its influence is compensated by counteracting factors with the result that no net effect remains. Consequently, the principle of relativity merely expresses operational limitations but does not extend to the nature of the concepts involved.

By contrast, Einstein understood the principle of relativity in a stronger, more literal sense. The results obtained by differently moved observers are objectively equivalent, not merely indistinguishable in their appropriateness. There is no true, universal measure of the relevant quantities; rather, electromagnetic fields and spatiotemporal relations are really different in different frames of reference. In Einstein's approach, distant simultaneity is a relational notion in that it is only defined with respect to a frame of reference. Frame dependence (or "relativity") is part of the concept of simultaneity, it is not merely an obstacle to the appropriate application of the concept. Einstein's insistence on the operational foundation of scientifically adequate concepts was accepted as a model by empiricist positions to the philosophy of science. Moritz Schlick, Hans Reichenbach, or Percy Bridgman regarded Einstein's emphasis on the definitional and procedural aspects of concept formation as a major breakthrough in epistemology.

The idea to elucidate the semantic features of a concept by drawing on the characteristics of the pertinent measurement procedures comes out particularly

clear in the claim of the conventionality of simultaneity. This claim originated with Poincaré [21, p. 13], was accepted by Einstein [11, p. 28] and elaborated within the mentioned empiricist approaches. As the argument developed in this latter framework runs, the evaluation of one-way velocities requires distant simultaneity relations. Yet in order to single out simultaneous events, signal speed needs to be known. As a result of this reciprocal dependence, distant simultaneity cannot be based on experience alone but is (within limits) subject to a stipulation ([25, pp. 148–149]; [27, p. 155]).

Underlying such commitments to operationalism is the conviction that establishing the concept of distant simultaneity requires a feasible method for comparing events in temporal respect. The crucial step is, then, that the room left by such procedures is indicative of the room inherent in the concept. On this markedly operational attitude, epistemic confines in ascertaining simultaneity relations are tantamount to the objective indeterminateness of these relations. It is precisely this attitude that made Einstein accept the counterintuitive, frame-dependent judgments about simultaneity relations as an adequate aspect of the concept of simultaneity.

In sum, Poincaré took the first step and advocated a procedural approach toward the notion of simultaneity, Einstein went one step further and advanced a procedural notion of simultaneity. Viewed along such lines, the ramifications of introducing a worldwide web of electrocoordinated clocks reached far up into the lofty realms of theoretical physics and philosophy of science.

6 Technological Problems, Technological Solutions, and Scientific Progress

The incipient career of special relativity theory places the fruitful interaction of technology and physics in the lime light. The early development of Einstein's thought shows that technology can be heuristically fruitful; it can promote scientific understanding. This finding does not square well with either one of the before-mentioned accounts of the relationship between pure and applied research. Neither the cascade model nor the emergentist conception left room for a seminal or productive influence of technology on science. The conclusion to be drawn from the case study is that technological challenges need not have a deteriorating effect on science. It may happen that technology stimulates scientific inventiveness.

At first sight, this account does not precisely respond to the concern raised earlier. Einstein had a scientific problem which he solved by developing a technology-based solution. The predicament addressed before was that focusing on technological problems might bring scientific progress to a halt. The worry mentioned in Sect. 2 was that concentrating research on technological issues could exhaust the epistemic resources of science and eventually block any further advancement of scientific understanding. However, the story of how special relativity was conceived can be reframed in such a way that concerns of this sort

are mitigated. After all, the technological problems of establishing simultaneity that plagued railroad planners, cartographers, and other practical professions stimulated Poincaré and Einstein to revise the conception of simultaneity. As a result of application pressure, they conceived of simultaneity as a definition and suggested an operational approach. At second glance, therefore, the account is able to alleviate apprehensions as to the harmful impact of applied research on epistemic science.[2]

In other cases the pressure of practical problems on theory development is even more pronounced. Not infrequently, practical challenges cannot appropriately be met without treating problems in basic science. This feature I call *application innovation*. It involves the emergence of theoretically significant novelties within the framework of use-oriented research projects. Although theoretical understanding is not among the objectives of applied research, it may yet be produced in the course of solving practical problems. On some occasions, treating such problems successfully demands addressing epistemically significant issues. Once in a while, applied research naturally grows into basic science and cannot help generating epistemic insights.

High-temperature superconductivity is a case in point. The phenomenon was discovered in 1986 in the IBM research laboratory near Zurich, and its identification stimulated the development of new theoretical accounts of superconductivity. Similarly, the transistor effect was found in the Bell laboratories. The emergence of this effect was based on the truly innovative procedure of adding impurities to semiconductors which act as electron donors or acceptors. This idea enriched solid state physics tremendously. Turning to biology, the path-breaking polymerase chain reaction was first conceived in a biotechnology firm, and the revolutionary conception of prions was elaborated in the practical context of identifying infectious agents. Prions are infectious proteins which reproduce without the assistance of nucleic acids; they were discovered during a use-oriented study on the sheep disease scrapie.

In these examples, research had been directed toward a practical goal but unintentionally produced innovations in basic research. This is no accident. Applied research tends to transcend applied questions for methodological reasons. A lack of deeper understanding of a phenomenon eventually impairs the prospects of its technological use. Superficial empirical relations, bereft of theoretical understanding, tend to collapse if additional factors intrude. Uncovering the relevant mechanisms and embedding them in a theoretical framework is of some use typically for ascertaining or improving the applicability of a finding. Scientific understanding makes generalizations robust in the sense that the limits of validity

[2] If the story is told in this way, the early development of quantum mechanics can be taken as a continuation. In his 1930 introduction to quantum theory, Werner Heisenberg explicitly placed his approach within the tradition of Einstein's operational analysis of seemingly innocuous concepts. As Heisenberg argued, what Einstein accomplished with respect to simultaneity, he aimed to do with respect to the concept of observation. Quantum theory needs to be based conceptually on the recognition that the interaction between object and observer can neither be neglected nor controlled [28, pp. 2–3].

can be anticipated or, as the case may be, expanded. Treating applied questions appropriately requires not treating them exclusively as applied questions. This is why epistemic science has less to fear from application pressure than it might appear initially.

The cascade model applies in a number of cases, and the emergentist approach rightly characterizes others. Yet application innovation represents a third mode of research which teaches a methodological lesson different from the others and tends to vindicate applied research in methodological respect.

The electrodynamics of moving bodies headed the research agenda of the period. Einstein approached this familiar problem situation in an unfamiliar way, namely, by starting from a procedural notion of simultaneity. This notion itself was by no means novel; it emerged tied up with the progress of clock technology and lay open right in front of Einstein's eyes at his desk at the patent office. The innovative step Einstein took was to connect topical areas and to bring the practice of railroad planners and surveyors to bear on issues of highbrow physical theory. This is quite typical of human creativity. On rare occasions only do we succeed in conceiving ideas completely novel and without precedent. Much more often innovations are produced by the more modest procedure of bringing together what appeared separate before. The Copernican achievement is precisely of this sort. Copernicus intended to solve the problem of the apparent inequality of planetary motions, as many had attempted before him, and he did so by drawing on the heliocentric ordering of the planetary orbits that was bequeathed to him by astronomical tradition. Both elements were widely known. Yet no one had endeavored before to invoke the heliocentric configuration as a means for resolving the inequality problem. Unifying seemingly disparate features is the predominant mode of producing innovations. And this is precisely the mechanism underlying Einstein's originality. He linked the technology of clock coordination to the issue of how bodies move when they approach the velocity of light. Links of this sort are the stuff human originality is made of.

References

1. F. Bacon: *Neues Organon I-II* (1620); Latin–German translation and edition by W. Krohn (Meiner, Hamburg 1999).
2. R. Descartes: *Discours de la méthode* (1637); German translation and edition by L. Gäbe (Meiner, Hamburg 1960).
3. I. Hacking: Representing and Intervening. Introductory Topics in the Philosophy of Natural Science (Cambridge University Press, Cambridge 1983).
4. V. Bush: *Science The Endless Frontier. A Report to the President* (United States Government Printing Office, Washington D.C. 1945); online: `http://www.nsf.gov/od/lpa/nsf50/vbush1945.htm`
5. R. Hohlfeld: Strategien gegen Krebs – Die Planung der Krebsforschung, **in** W. van den Daele, W. Krohn & P. Weingart (eds.): *Geplante Forschung. Vergleichende Studien ber den Einfluß politischer Programme auf die Wissenschaftsentwicklung* (Suhrkamp, Frankfurt 1979), p. 181.
6. P. Galison: *Einstein's Clocks, Poincaré's Maps. Empires of Time* (W.W. Norton, New York 2003).

7. N. Cartwright: Fundamentalism versus the Patchwork of Laws, **in** D. Papineau (ed.): *The Philosophy of Science* (Oxford University Press, Oxford 1996), p. 314.
8. T. Wilholt: Design Rules: Industrial Research and Epistemic Merit, *forthcoming* (2005).
9. J. Ziman: The Continuing Need for Disinterested Research, *Science and Engineering Ethics* **8**, 397 (2002).
10. S.S. Schweber: Physics, Community and the Crisis in Physical Theory, *Physics Today*, November 1993, p. 34.
11. A. Einstein: Zur Elektrodynamik bewegter Körper, *Ann. Physik* **17**, 891 (1905); reprinted in [29, pp. 26–50].
12. H.A. Lorentz: Simplified Theory of Electrical and Optical Phenomena in Moving Systems, **in** [16], p. 255, reprinted from 1899.
13. P. Drude: *The Theory of Optics* (Dover, New York 1959), translated from *Lehrbuch der Optik* (S. Hirzel, Leipzig 1900).
14. H. Poincaré: La mécanique nouvelle, **in** É. Guillaume (ed.): *La Mécanique Nouvelle* (Gauthier-Villars, Paris 1927), 1-17. Reprint from (1909).
15. R. McCormmach: Einstein, Lorentz, and the Electron Theory, *Historical Studies in the Physical Sciences* **2**, 41 (1970).
16. K. Schaffner: *Nineteenth–Century Aether Theories* (Pergamon Press, Oxford 1972).
17. M. Carrier: Shifting Symbolic Structures and Changing Theories: On the Non-Translatability and Empirical Comparability of Incommensurable Theories, **in** M. Ferrari & I. Stamatescu (eds.): *Symbol and Physical Knowledge. On the Conceptual Structure of Physics* (Springer, Berlin 2001), pp. 125-148.
18. M. Carrier: Semantic Incommensurability and Empirical Comparability: The Case of Lorentz and Einstein, *Philosophia Scientiae* **8**, 73 (2004).
19. H.A. Lorentz: Electromagnetic phenomena in a system moving with any velocity less than that of light, *Proc. Acad. Science Amsterdam* **IV**, 669 (1904); reprinted in [29].
20. K. Schaffner: Einstein versus Lorentz: Research Programmes and the Logic of Comparative Theory Evaluation, *The British Journal for the Philosophy of Science* **25**, 45 (1974).
21. H. Poincaré: La mesure du temps, *Revue de métaphysique et de morale* **6**, 1 (1898).
22. H. Poincaré: La théorie de Lorentz et le principe de réaction, **in** *Œuvres de Henri Poincaré IX* (Gauthier-Villars, Paris 1954), pp. 464-488.
23. H. Poincaré: *La Valeur de la Science* (Flammarion, Paris 1905).
24. A.I. Miller: *Albert Einstein's Special Theory of Relativity. Emergence (1905) and Early Interpretation (1905-1911)* (Addison–Wesley, Reading Mass. 1981).
25. H. Reichenbach: *The Philosophy of Space and Time* (Dover Publications, New York 1958); translated from *Philosophie der Raum–Zeit–Lehre* (W. de Gruyter, Berlin 1928).
26. J. Messerli: *Gleichmässig, pünktlich, schnell. Zeiteinteilung und Zeitgebrauch in der Schweiz im 19. Jahrhundert* (Chronos, Zürich 1995).
27. M. Carrier: *The Completeness of Scientific Theories. On the Derivation of Empirical Indicators within a Theoretical Framework: The Case of Physical Geometry*, Western Ontario Series in the Philosophy of Science 53 (Kluwer Academic Publishers, Dordrecht 1994).
28. W. Heisenberg: *Die Physikalische Prinzipien der Quantentheorie* (B.I. Wissenschaftsverlag, Mannheim 1958); English translation: *The Physical Principles of the Quantum Theory* (Dover Publications, New York 1949).
29. H.A. Lorentz, A. Einstein & H. Minkowski: *Das Relativitätsprinzip. Eine Sammlung von Abhandlungen* (Wissenschaftliche Buchgesellschaft, Darmstadt 1923).

Part II

Foundation and Formalism

Foundations of Special Relativity Theory

J. Ehlers

Max Planck Institute for Gravitational Physics (Albert–Einstein–Institute),
Am Mühlenberg 1, 14476 Golm, Germany
juergen.ehlers@aei.mpg.de

1 Introduction

Any physical theory is based partly on a spacetime structure, which is needed to locate events (= spacetime points) and to provide a domain of definition for variables describing particles and fields.

So far, the following spacetime structures have been successfully used in physics: (i) Newton's spacetime, (ii) the Einstein–Minkowski spacetime of Special Relativity (SR) and (iii) the Riemann–Einstein spacetime of General Relativity (GR). In the first two, metric and (flat) connection are specified once and for all, influencing matter but not affected by matter. In GR the metric and the associated connection and curvature are physical fields. Accordingly, in Newtonian and SR physics one can separate kinematics from dynamics, which is not possible in GR.

The metric not only serves to determine distances, time spans and causal relations, it always enters the description of matter. This fact is often obscured by the habit of not distinguishing between vectors and covectors (= one forms) using the metric to "move indices". While phases of waves are represented geometrically by hypersurfaces given infinitesimally by $k_\alpha = \partial_\alpha S$, world lines or rays are curves, i.e., generated by vectors p^α (e.g., kinetic 4–momentum). To relate particle and wave quantities à la Einstein–de Broglie, one writes

$$p^\alpha = \hbar g^{\alpha\beta} k_\beta \quad \text{or} \quad \hbar k_\alpha = g_{\alpha\beta} p^\beta \,, \tag{1}$$

or one relates the canonical momentum to the 4–velocity via

$$\overline{p}_\alpha = m g_{\alpha\beta} u^\beta \,, \quad u^\alpha = \frac{dx^\alpha}{ds}. \tag{2}$$

In my view, these equations together with the interpretations of p^α, k_α, $\overline{p}_\alpha$, and u^α and similar relations between vectors and covectors exhibit the basic role of the spacetime metric for physics as well as statements about distances or time intervals. In other words, the basic role of the metric is to define an isomorphism between the tangent space and its dual. Once this is given, other roles of the metric (distance, duration, angle) can be derived.

J. Ehlers: *Foundations of Special Relativity Theory*, Lect. Notes Phys. **702**, 35–44 (2006)
DOI 10.1007/3-540-34523-X_3
© Springer-Verlag Berlin Heidelberg 2006

2 Inertial Frames

To start physics one needs a frame of reference, a way to locate events by means of coordinates. In Newtonian physics as well as in SR it is assumed that there are preferred frames, inertial frames, with coordinates $x^\alpha = (x^a, x^0)$, $x^0 = ct$, where the x^a are rectangular Cartesian coordinates in position space which is taken to be Euclidian, and t measures time. Taking distance and duration as basic measurable quantities presupposes the existence of reproducible standards of length and time functioning independent of their pre–histories.

While in Newtonian mechanics an absolute time t is assumed and inertial frames are characterized by the law of inertia $\ddot{x}^a = 0$, in SR one proceeds differently, for well–known reasons. One assumes, as laws of nature,

1. that in empty space light propagates independently of the state of motion of the source, and
2. relative to an inertial frame, the mean speed of light on any closed triangular path in vacuo has a universal value c.

Combined with Einstein's definition of simultaneity of pairs of events, these two properties imply the existence of a global time coordinate t in an inertial frame such that the one–way speed of light in vacuo is always c. (The last assertion involves a *convention* on simultaneity and *facts* about light propagation.) One may then also assume, relative to an SR inertial frame, the law of inertia for "free" particles as stated above, but now with respect to the Einsteinian time t. In inertial coordinates, light signals are then given by

$$\Delta\mathbf{x}^2 - (\Delta x^0)^2 \equiv \eta_{\alpha\beta}\Delta x^\alpha \Delta x^\beta = 0 \,, \tag{3}$$

and free particles by

$$\frac{d^2x^\alpha}{dt^2} = 0 \,. \tag{4}$$

Statement 1. does not require the concept of speed for either light or sources. It says that the set of events constituing a short light pulse is uniquely determined by the emission event. It holds equally in GR.

Because of assumption 1., in 2. the state of motion of the light source does not matter provided reflecting mirrors (tacitly assumed in 2.) are considered as sources of the reflected light.

Historical remarks. The statement 1. was originally taken from Maxwell's theory, unambiguous experimental support came only in the 1960ies, e.g., [1, 2]. In 1905 and later, Einstein assumed that his definition of simultaneity is an equivalence relation between events. In 1922 H. Weyl replaced Einstein's assumption by the experimentally testable statement 2.

3 Poincaré Transformations

Suppose (x^a) and $(x^{a'})$ are two inertial coordinate systems I, I'. It is assumed that events can be identified as being "the same" from both systems. Then there

is a bijective transformation $x^\alpha \mapsto x^{\alpha'}$. There are several ways to determine its form, based on different assumptions.

I first present explicitly an elementary derivation. It presupposes (3) and (4) to hold in both frames, for unspecified coordinate ranges. Take then two free particles P_1, P_2 which have in I the same 3–velocity. Imagine light signals to be sent back and forth between P_1 and P_2. This defines, according to (3) and (4), a sequence of events on P_1 which have equal time separations Δt (light clock). The assumptions (1) and (2) about light rays and free particles imply that, viewed from I' the particles P_1 and P_2 move with constant 3–velocities, and light rays move back and forth between them. We next show that the 3–velocities of P_1 and P_2 in I' are equal, too. For this purpose we interpret – separately in I and I' – the coordinates x^α and $x^{\alpha'}$ as affine coordinates of two distinct 4–spaces S, S'. (This is necessary since at this stage we do not yet have a spacetime geometry). Then, the free particle motions and the light rays are represented, according to (3) and (4), as straight lines both in S and S'. In S, by construction, P_1 and P_2 are parallel, and so are the two sets of light ray segments connecting them. The crucial point now is that parallelism of two free particle world lines can be expressed solely in terms of light rays and a few auxiliary free particle world lines as recognized by Marzke [9], see Fig. 1. Therefore, P_1 and P_2 are represented by parallel lines also in S', and consequently, viewed from I', we also have a light clock with equal "ticks" $\Delta t'$. And since P_1 and P_2 can be chosen arbitrarily close to each other, the "periods" Δt, $\Delta t'$ can be made arbitrary small, and therefore the time coordinates t, t' on P_1 are related by an equation $t' = at + b$ with constants a, b with $a > 0$ with natural time orientation. Therefore, the transformation $x^\alpha \mapsto x^{\alpha'}$ must be such that

$$x^{\alpha'}(t) = f^{\alpha'}(x_0^\beta + tc^\beta) = x_0^{\alpha'} + tc^{\alpha'} \tag{5}$$

holds for arbitrary x_0^β, c^β (with some open ranges). Note that the independent variable on both sides is t, due to the preceding argument. Differentiating twice w.r.t. t gives that the transformation is affine,

$$x^{\alpha'} = A^{\alpha'}_\beta x^\beta + d^{\alpha'}. \tag{6}$$

Using next the invariance of (3) shows that the matrix A must be a positive multiple λL of a Lorentz matrix,

$$A = \lambda L\ , \quad \lambda > 0\ , \quad L \in O(3,1). \tag{7}$$

In fact, (3) says that the quadratic form

$$\Delta x^{\alpha'} \eta_{\alpha'\beta'} \Delta x^{\beta'} = A^{\alpha'}_\beta \eta_{\alpha'\gamma'} A^{\gamma'}_\delta \Delta x^\delta \Delta x^\beta \tag{8}$$

vanishes if and only if $\eta_{\alpha\beta}\Delta x^\alpha \Delta x^\beta$ does, where

$$(\eta_{\alpha\beta}) = (\eta_{\alpha'\beta'}) := \mathrm{diag}(1,1,1,-1)\ . \tag{9}$$

This holds if and only if

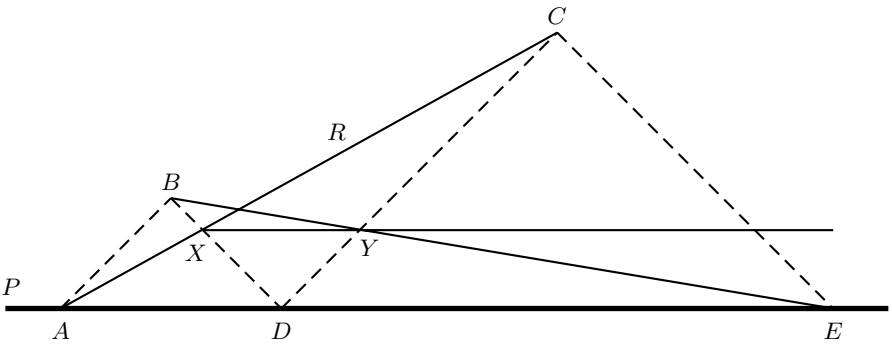


Fig. 1. *Task:* Given the (particle) line P and the event $X \notin P$, construct the parallel to P through X. *Construction:* Choose A on P, draw $AX = R$. Draw light ray DX and light ray from A that meets DX in B. Draw light ray through D parallel to AB to obtain C. Draw a light ray through C parallel to BD to obtain E. Draw BE to obtain Y. Then $XY \parallel P$. Note that the whole construction takes place in a plane of S, and therefore – because of (3) and (4) – in a plane of S'

$$A_\beta^{\alpha'} \eta_{\alpha'\gamma'} A_\delta^{\gamma'} = \lambda^2 \eta_{\beta\delta} \quad \text{with} \quad \lambda > 0 \,, \tag{10}$$

which means that $\lambda^{-1} A = L$ is a Lorentz metric. Note that the invariance of (3) follows from property 1. of Sect. 2. As long as statements about a single inertial frame are made, only property 2. about light is needed; for synchronizing clocks one may use light signals emitted by sources at rest in the frame considered, and (3) holds for those signals – separately in each such frame. However, because of 1. the sequence of events of which a light ray consists is the same in all inertial frames, and this fact implies the invariance of (3). (Both (3) and (4) establish "bridges" between I and I'.) It follows (!) that one may, without loss of generality, take the coordinate ranges in both frames to be $\mathbb{R}^4$. This is as far as one can get under the specified assumptions.

One can put $\lambda = 1$ if one requires that in I and I' the same units of length and time are employed, e.g. by defining the (proper) time via a particular Cs transition and fixing the value of c, as agreed since 1983. To infer $\lambda = 1$ from that requires however, an additional assumption, e.g. that time–dilation be symmetric between I and I', or that meter sticks at rest in I and I', orthogonal to the direction of relative motion, coincide when passing each other.

Another way to deduce (6), (7), which is more elegant than the above one, but perhaps less physical, takes for granted only that the coordinate range is $\mathbb{R}^4$ in any inertial frame and that the relation (3) between event pairs is invariant. The statement then is:

A bijection $\mathbb{R}^4 \mapsto \mathbb{R}^4$ is of the form (6), (7) if and only if (3) is invariant as has been shown by Borchers and Hegerfeld [10].

Note that, in contrast to the previous argument, no differentiability is assumed, and nothing is assumed about particle motions. Given $\mathbb{R}^4$ as the space-

time manifold, the light cone defines the kinematics of SR (except for a common spacetime scale).

Remarks. The affine nature of the transformations between SR inertial frames, (6), has often been *taken for granted*, or as obvious, or based on vaguely stated arguments (as in Einstein 1905). It has been *derived* by H. Weyl, A.D. Alexandrov, G. C. Zeeman and others. – *Assuming* linearity, Robertson [3] has derived (7) from idealized inferences from the classic optical experiments by Michelson–Morley [4], Kennedy–Thorndike [5], and Ives–Stilwell [6]. – As shown above, rods and clocks are not needed, free particles and light rays are sufficient to obtain the SR kinematics.

4 Minkowski Spacetime

The considerations reviewed in Sects. 2 and 3 can be restated, following H. Minkowski (1908), as follows:

Spacetime can be represented geometrically as the (real), 4-dimensional affine space A^4, equipped with a Lorentz metric $\eta_{\alpha\beta}$ (here taken to have signature $+++-$) compatible with the affine structure. In this interpretation, an inertial coordinate system is an affine coordinate system, (pseudo-) orthonormal with respect to the metric $\eta_{\alpha\beta}$; it is given by an Origin O and an orthonormal basis (E_α) of vectors, and then an event P has coordinates x^α, $\overrightarrow{OP} = x^\alpha E_\alpha$.

Using Minkowskian geometry it is easy to derive the "special" Lorentz transformation

$$x^{1'} = \gamma(v)(x^1 - \beta x^0)\ , \quad x^{2'} = x^2, \quad x^{3'} = x^3 \tag{11}$$

$$x^{0'} = \gamma(v)(x^0 - \beta x^1)\ . \tag{12}$$

Indeed, given two inertial frames with bases (E_α), $(E_{\alpha'})$ and, without loss of generality, the same origin O, one can employ spatial rotations to arrange that E_1 and E_1' are contained in the (timelike) plane spanned by E_0 and $E_{0'}$, and $E_2 = E_{2'}$, $E_3 = E_{3'}$ in the (spacelike) plane orthogonal to the $(E_0, E_{0'})$–plane. This results in the standard equations written down above. (Without Minkowskian geometry it is not easy to derive those equations since the 3-spaces of the two frames "move" relative to each other.) The reference space of an inertial frame I consists of parallel, timelike lines; events which are simultaneous in I fill a spacelike hypersurface orthogonal to those lines.

The light cone of an event separates vectors X^α into timelike ones ($\eta_{\alpha\beta}X^\alpha X^\beta = X\cdot X < 0$), lightlike ones ($X^2 = 0$, $X \neq 0$) and spacelike ones ($X^2 > 0$ or $X = 0$).

The Poincaré transformations (6), (7) with $\lambda = 1$ relate inertial coordinates. Since they are characterized by leaving certain relations invariant they form a group. This group may also be taken to act ("actively") on A^4; it is the isometry group of (A^4, η); it acts simply transitively on the set of inertial coordinate systems.

Due to the foregoing, tensors (and, if needed, spinors) may be used to describe physical objects in spacetime.

5 Axiomatics

Minkowski spacetime has been axiomatized in the style of Hilbert's axiomatization of Euclidean geometry, even in several versions.

1. The oldest axiomatization has been elaborated by A.A. Robb [11]; he uses as the only primitive relation between events the causal order, $p \gg q$, p causes q.
2. C. Caratheodory [12] and H. Reichenbach [13] used particles and light rays (or "first signals") and coincidence relations.
3. B. Mundy [14] gave two axiomatizations, one based on optical connectivity, the other one on several betweenness-relations on timelike and lightlike lines.
4. H. Weyl [15] sketched an axiomatic, using free particles and light rays.
5. J. Schutz [16] also used free particles and light rays as primitive concepts. His treatment is mathematically remarkable in that he proved the independence of all his axioms.

This list is not complete, I just wanted to mention that the geometry of Minkowski spacetime has been axiomatized in all rigour. Hence, skeptics may be assured that SR as a logico–mathematical theory is as free of contradictions as Euclid's geometry in Hilbert's version.

6 The Principle of Special Relativity and Its Limits

By definition, inertial frames are equivalent as far as free particle motions, light rays and statements about measurements of lengths, times, frequencies on moving bodies, light sources etc. are concerned.

Einstein's *principle of SR* postulates that this equivalence holds for all laws of physics. This means that the isometries of Minkowski spacetime are taken to induce symmetries in all physical laws.

If laws can be expressed in terms of tensor (or spinor) calculus they satisfy that principle automatically, and for this reason it has been possible to formulate, or reformulate, large parts of physics within the framework of Poincaré invariant, special theory, and these branches of physics turned out to be empirically successful. However, SR met with difficulties in two ways. One is that, as Einstein showed, gravitation does not fit naturally into SR. The flat metric has to be replaced by a curved one and becomes a dynamical field, and from the point of view of the resulting general theory of relativity, SR is relegated to an approximation. The other, deeper limitation is that relativistic quantum theories require a radical change of the microstructure of spacetime; substitutes for a manifold are under discussion, but the problem seems to be far from solved. In the first respect, special relativity did not survive the previous century, in the second respect, even its large-scale successor GR will not survive the 21^{st} century. The question of this meeting can perhaps be rephrased as: Will SR survive as a tool for perturbative analyses of certain classes of experiments?

Additional remark. In the change SR → GR the metric changes from a pre–assigned, absolute element into a dynamical field with its own degrees of freedom. While both SR and GR can be formulated in a generally covariant form, the Poincaré group loses its role as invariance group in the transition SR → GR. (For a recent discussion of this aspect see, e.g. N. Straumann [17].)

7 Examples

As examples of SR laws I recall Maxwell's equations of electrodynamics,

$$F_{[\alpha\beta,\gamma]} = 0 \ , \quad F^{\alpha\beta}{}_{,\beta} = J^{\alpha} \tag{13}$$

and "Euler's" equations for perfect fluids,

$$T^{\alpha\beta}{}_{,\beta} = \left[(\rho + p)U^{\alpha}U^{\beta} + pg^{\alpha\beta}\right]_{,\beta} = 0 \tag{14}$$

with $\rho = f(p) > 0$, $0 \leq f' \leq 1$. Both systems separately, with $J^{\alpha} = 0$ in (13), admit well–posed initial value problems, and (13) implies stress–energy–momentum conservation. If the fluid represents a plasma, the systems may be coupled via J^{α}, then the total energy tensor is conserved.

In empty space the characteristics of (13) are null hypersurfaces and the bicharacteristics are light–like straight lines. WKB approximation can be used to show that the initial assumption about light rays now follows from theory.

In a similar way one can derive the law of inertia (4) from $T^{\alpha\beta}{}_{,\beta} = 0$, applied to an isolated, finite body. If $-T^{\alpha}{}_{\beta}$ maps the interior of the future light cone into itself, which is one version of an energy condition, then the convex hull of the body contains a timelike straight line, its center–of–mass line, so that a small body can be represented approximately by such a line given by (4), see J.L. Synge [18].

This and the preceeding remark show that the simple assumptions made to initiate the theory later follow from the laws within the theory, a requirement sometimes called *semantic consistency.*

8 Accelerated Frames of Reference

For recognizing inertial forces and as stepping stones leading to GR it is useful to consider accelerated frames of reference in flat spacetime.

Applying the transformation

$$\begin{aligned} x &= \left(\frac{c^2}{g} + X\right)\cosh\left(\frac{g}{c}T\right) \\ ct &= \left(\frac{c^2}{g} + X\right)\sinh\left(\frac{g}{c}T\right) \\ y &= Y \ , \ z = Z \ , \ \mathbf{X} = (X, Y, Z) \end{aligned} \tag{15}$$

to the flat line element

$$ds^2 = d\mathbf{x}^2 - c^2 dt^2 \tag{16}$$

gives

$$ds^2 = d\mathbf{X}^2 - \left(1 + \frac{gX}{c^2}\right)^2 c^2 dT^2 \,, \tag{17}$$

with $-c^2/g < X < \infty$. A point $\mathbf{X}$ of the frame has the absolute dynamical acceleration $g/(1+gX/c^2)$. Equation (17) exhibits the time dilation (measured by Pound–Rebka–Snider [7,8]) and, via $ds^2 = 0$, gives an effective index of refraction $n = \left(1 + gX/c^2\right)^{-1}$, which leads to curved light paths in the accelerated 3–space. Since gX plays the part of gravitational potential, Einstein conceived in 1907 [22] (by sophisticated arguments – before Minkowski's geometry), guided by his equivalence principle, the ideas of gravitational time dilation and light bending, in a gravitational field with potential Φ in place of gX.

Similarly, using cylindrical coordinates and substituting $\varphi \mapsto \varphi - \omega t$, one gets, for $\omega r < c$,

$$ds^2 = dr^2 + \left(\frac{r d\varphi}{\sqrt{1 - (\frac{\omega r}{c^2})^2}}\right)^2 + dz^2 - \left(1 - \left(\frac{\omega r}{c}\right)^2\right)\left(cdt - \frac{\frac{\omega r}{c}}{1 - (\frac{\omega r}{c})^2} r d\varphi\right)^2 \,, \tag{18}$$

the metric in a uniformly rotating frame. It exhibits the time delay related to the centrifugal potential, the Coriolis (vector) potential which affects simultaneity in the rotating frame, and the Lorentz contraction which "causes" the space–geometry to be non–Euclidean.

The use of curvilinear coordinates and/or non–parallel orthonormal bases (tetrads) in SR does, of course, not affect the intrinsic Poincaré invariance of SR, nor does it introduce "true" gravitational tidal fields, i.e. curvature.

Both examples (17) and (18) exhibit rigid motions in Born's sense, while rigid bodies do not exist in SR. (In Minkowski spacetime, all rigid motions are known due to G. Herglotz [19] and F. Noether [20].)

9 SR Causality

The laws of SR are local ones. Poincaré invariance by itself does not exclude signals with $v > c$, nor does it forbid characteristics outside the light cone. However, as Einstein remarked in 1907, if such signals were assumed to exist, an observer could send a message which would arrive at the receiving station earlier than it was sent, judged by the inertial time of the observer. Although this does not entail a logical contradiction, Einstein remarks, it contradicts all our experience such that its impossibility seems to be established sufficiently.

Since Einstein used his – after all, conventional – concept of simultaneity, one may not be convinced. However, if there were superluminal light signals obeying Poincaré invariant laws, and if observers were able to trigger such signals, then they should be able to affect their own past. Therefore, if one wants to maintain,

as part of the idea of causality, that the past relative to any observer is factual and cannot be changed by any actions, then one is forced to postulate (!) that causal influences can propagate with speeds at most equal to c, and hence that partial differential equations expressing laws of nature must be hyperbolic, with characteristic ray cones inside or coinciding with the light cone.

I find it interesting to note that Heisenberg [21] draws a parallel to that "postulate" and his assumption that quantum mechanics requires "uncontrolled disturbances" during measurements. In both cases the assertions do not *follow* from laws of the theory, but they have to be *postulated* in order that we can maintain our usual thinking about causation, past and future and not get into contradictions with the formal theories used to describe phenomena, SR and quantum mechanics, respectively.

This, then leads to *Einstein causality*: The causal past of an event P consists of its past half null cone, its domain of influence consists of its future half null cone. This causal structure is maintained in GR as well as in local quantum field theory. A substitute for it in a not–yet existing quantum gravity theory or TOE is not known, though there are attempts to find one.

References

1. T. Alväger, F.J.M. Farley, J. Kjellmann, and I. Wallin: Test of the second postulate of Special Relativity in the GeV region, *Phys. Lett.* **12**, 260 (1964).
2. K. Brecher: Is the speed of light independent of the velocity of the source?, *Phys. Rev. Lett.* **39**, 1051 (1977).
3. H.P. Robertson: Postulate versus observation in the Special Theory of Relativity, *Rev. Mod. Phys.* **21**, 378 (1949).
4. A.A. Michelson and E.W. Morley: On the relative motion of the Earth and the luminiferous ether, *Am. J. Sci.* **34**, 333 (1887).
5. R.J. Kennedy and E.M. Thorndike: Experimental establishment of the relativity of time, *Phys. Rev.* **42**, 400 (1932).
6. H.E. Ives and G.R. Stilwell: An experimental study of the rate of a moving clock, *J. Opt. Soc. Am.* **28**, 215 (1938).
7. R.V. Pound and G.A. Rebka, G.A.: Apparent weight of Photons, *Phys. Rev. Lett.* **4**, 337 (1960).
8. R.V. Pound and J.L. Snider: Effect of gravity on gamma radiation, *Phys. Rev.* **140**, B788 (1965).
9. R.F. Marzke, see M. and Wheeler, ch. 3 in *Gravitation and Relativity*, H.–G. Chin et al (eds.), Benjamin Inc., N.Y. 1964, pp. 50-52.
10. H.–J. Borchers and G.C. Hegerfeld, *Commun. Math. Phys.* **28**, 259 (1972).
11. A.A. Robb: *A Geometry of Time and Space*, Cambridge 1914, later editions 1921, 1936 ibid.
12. C. Caratheodory, Zur Axiomatik der speziellen Relativitätstheorie, *Sitzungsber. Preuss. Akad. Wissensch.*, 1924, S. 12
13. H. Reichenbach, *Philosophie der Raum–Zeit–Lehre*, 1928 (English translation: *The philosophy of space and time*, Dover Publications, New York, 1958).
14. B. Mundy, The physical content of Minkowski geometry, *British Journal of Philosophy of Science* **37**, 25 (1986); Optical Axiomatization of Minkowski Space–Time Geometry, *Philosophy of Science* **53**, 1 (1986).

15. H. Weyl, Kap II. u. Anhang aus der Vorlesung "Axiomatik", Göttingen 1930/31.
16. J.W. Schutz, Independent Axioms for Minkowski Space-Time, Pitman Research Notes in Mathematics Series 373 (Longman, Harlow, 1997). This book contains many references on SR axiomatics.
17. N. Straumann, *General Relativity*, Springer, Berlin 2004.
18. J.L. Synge, Relativity: The Special Theory, Amsterdam 1956, esp. ch. VIII, B' 16.
19. G. Herglotz, Über den vom Standpunkt des Relativitätsprinzips aus als 'starr' zu bezeichnenden Körper, *Ann. Physik* **31**, 393 (1910).
20. F. Noether, Zur Kinematik des starren Körpers in der Relativitätstheorie, *Ann. Physik* **31**, 919 (1910).
21. W. Heisenberg, *Die physikalischen Prinzipien der Quantentheorie*, Leipzig 1930, insbes. Einleitung.
22. A. Einstein, Über die vom Relativitätsprinzip geforderte Trägheit der Energie, *Ann. Physik* **23**, 371 (1907).

Algebraic and Geometric Structures in Special Relativity

D. Giulini

Physikalisches Institut der Universität Freiburg Hermann–Herder–Straße 3, 79104 Freiburg, Germany
giulini@physik.uni-freiburg.de

Abstract. I review, on an advanced level, some of the algebraic and geometric structures that underlie the theory of Special Relativity. This includes a discussion of relativity as a symmetry principle, derivations of the Lorentz group, its composition law, its Lie algebra, comparison with the Galilei group, Einstein synchronization, the lattice of causally and chronologically complete regions in Minkowski space, rigid motion, and the geometry of rotating reference frames. Representation-theoretic aspects of the Lorentz group are not included. A series of appendices present some related mathematical material.

1 Introduction

In this contribution I wish to discuss some structural aspects of Special Relativity (henceforth abbreviated SR) which are, technically speaking, of a more advanced nature. Most of what follows is well known, though generally not included in standard text-book presentations. Against my original intention, I decided to not include those parts that relate to the representation- and field-theoretic aspects of the homogeneous and inhomogeneous Lorentz group, but rather to be more explicit on those topics now covered. Some of the abandoned material is (rather informally) discussed in [26]. All of it will appear in [22]. For a comprehensive discussion of many field-theoretic aspects, see e.g. [40].

I always felt that Special Relativity deserves more attention than what is usually granted to it in courses on mechanics or electrodynamics. There is a fair amount of interesting algebraic structure that characterizes the transition between the Galilei and Lorentz group, and likewise there is some interesting geometry involved in the transition between Newtonian (or Galilean) spacetime and Minkowski space. The latter has a rich geometric structure, notwithstanding the fact that, from a general relativistic viewpoint, it is "just" flat spacetime. I hope that my contribution will substantiate these claims. For the convenience of some interested readers I have included several mathematical appendices with background material that, according to my experience, is considered helpful being spelled out in some detail.

D. Giulini: *Algebraic and Geometric Structures in Special Relativity*, Lect. Notes Phys. **702**, 45–111 (2006)
DOI 10.1007/3-540-34523-X_4 © Springer-Verlag Berlin Heidelberg 2006

2 Some Remarks on "Symmetry" and "Covariance"

For the purpose of this presentation I regard SR as the (mathematical) theory of how to correctly implement the *Galilean Relativity Principle* – henceforth simply abbreviated by RP. The RP is a physical statement concerning a subclass of phenomena – those not involving gravity –, which translates into a mathematical symmetry requirement for the laws describing them. But there is no unique way to proceed; several choices need to be made, whose correctness cannot be decided by mere logic.

Given that the symmetry requirement is implemented by a group action (which may be relaxed; compare e.g. supersymmetry, which is not based on a group), the most fundamental question is: *what group?* In this regard there is quite a convincing string of arguments that, given certain mild technical assumptions, the RP selects either the Galilei or the Lorentz group (the latter for some yet undetermined velocity parameter c). This will be discussed in detail in Sect. 3.

Almost as important as the selection of a group is the question of how it should act on physical entities in question, like particles and fields. The importance and subtlety of this question is usually underestimated. Let us therefore dwell a little on it.

As an example we consider vacuum electrodynamics. Here the mathematical objects that represent physical reality are two spacetime dependent fields, $\mathbf{E}(\mathbf{x}, t)$ and $\mathbf{B}(\mathbf{x}, t)$, which take values in a vector space isomorphic to $\mathbb{R}^3$. There will be certain technical requirements on these fields, e.g. concerning differentiability and fall-off at spatial infinity, which we do not need to spell out here. For simplicity we shall assume that the set of all fields obeying these conditions forms an infinite-dimensional linear space $\mathcal{K}$, which is sometimes called the space of "kinematical" (or "kinematically possible") fields. Those fields in $\mathcal{K}$ which satisfy Maxwell's (vacuum) equations form a proper subset, $\mathcal{S} \subset \mathcal{K}$, which, due to the linearity of the equations and the boundary conditions (fall-off to zero value, say), is a linear subspace. It is called the space of "physical" (or "dynamically possible") fields. Clearly these notions of the spaces of kinematically and dynamically possible fields apply to all sorts of situations in physics where one considers "equations of motion", though in general neither of these sets will be a vector space. This terminology was introduced in [2].

In general, we say that a group G is a *symmetry group* of a given dynamical theory if the following two conditions are satisfied:

1. There exists an (say left-) effective action $G \times \mathcal{K} \to \mathcal{K}$, $(g, k) \mapsto g \cdot k$, of G on $\mathcal{K}$ (cf. Sect. A.1). Posing effectiveness just means that we do not wish to allow trivial enlargements of the group by elements that do not move anything. It also means no loss of generality, since any action of a group G on a set $\mathcal{K}$ factors through an effective action of G/G' on $\mathcal{K}$, where G' is the normal subgroup of trivially acting elements (cf. Sect. A.1. Such an action of the group on the kinematical space of physical fields is also called an *implementation* of the group into the physical theory.

2. The action of G on $\mathcal{K}$ leaves $\mathcal{S}$ invariant (as a set, not necessarily pointwise), i.e. if $s \in \mathcal{S}$ then $g \cdot s \in \mathcal{S}$ for all g. This merely says that the group action restricts from $\mathcal{K}$ to $\mathcal{S}$. Note that, from an abstract point of view, this is the precise statement of the phrase 'leaving the field equations invariant', since the field equations are nothing but a characterization of the subset $\mathcal{S} \subset \mathcal{K}$.

If this were all there is to require for a group to count as a *symmetry group*, then we would probably be surprised by the wealth of symmetries in Nature. For example, in the specific case at hand, we often hear or read the statement that the Lorentz group leaves Maxwell's equations invariant, whereas the Galilei group does not. Is this really true? Has anyone really shown in this context that the Galilei group cannot effectively act on $\mathcal{K}$ so as to leave $\mathcal{S}$ invariant? Certainly not, because such an action is actually known to exist; see e.g. Chap. 5.9 in [21]. Hence, in the general sense above, the Galilei group *is* a symmetry group of Maxwell's equations!

The folklore statement just alluded to can, however, be turned into a true statement if a decisive restriction for the action is added, namely that it be *local*. This means that the action on the space of fields is such that the value of the transformed field at the transformed spacetime point depends only on the value of the untransformed field at the untransformed point and *not*, in addition, on its derivatives.[1] This is the crucial assumption that is implicit in all proofs of Galilean-non-invariance of Maxwell's equations, and that is also made regarding the Lorentz group in classical and quantum field theories. The action of the Galilei group that makes it a symmetry group for Maxwell's equations is, in fact, such that the transformed field depends linearly on the original field and its derivatives to *all* orders. That is, it is highly non local.

Returning to the general discussion, we now consider a classical field, that is, a map $\Psi : M \to V$ from spacetime M into a vector space V. A spacetime symmetry-group has an action on M, denoted by $T : (g, x) \mapsto T_g(x)$, as well as an action on V, which in most cases of interest is a linear representation $g \mapsto D(g)$ of G. A local action of G on field space is then given by

$$(g, \Psi) \mapsto g \cdot \Psi := D(g) \circ \Psi \circ T_g^{-1}\,, \tag{1}$$

where here and below the symbol $\circ$ denotes composition of maps. This is the form of the action one usually assumes. Existing generalizations concerning possible non-linear target spaces for Ψ and/or making Ψ a section in a bundle, rather than just a global function, do not influence the locality aspect emphasized here and will be ignored.

Next to fields one also considers particles, at least in the classical theory. Structureless (e.g. no spin) particles in spacetime are mathematically idealized by maps $\gamma : \mathbb{R} \to M$, where $\mathbb{R}$ (or a subinterval thereof) represents parameter space. The parameterization usually does not matter, except for time orientation,

[1] Here one should actually distinguish between "ultralocality", meaning not involving any derivatives, and "locality", meaning just depending on derivatives of at most finite order.

so any reparameterization $f : \mathbb{R} \to \mathbb{R}$ with $f' > 0$ gives a reparameterized curve $\gamma' := \gamma \circ f$ which is just as good. On the space of particles, the group G acts as follows:

$$(g, \gamma) \mapsto g \cdot \gamma := T_g \circ \gamma \,. \tag{2}$$

Together (1) and (2) define an action on all the dynamical entities, that is particles and fields, which we collectively denote by the symbol Φ. The given action of G on that space is simply denoted by $(g, \Phi) \mapsto g \cdot \Phi$.

Now, the set of equations of motion for the whole system can be written in the general form

$$\mathcal{E}[\Sigma, \Phi] = 0 \,, \tag{3}$$

where this should be read as a multi-component equation (with 0 being the zero "vector" in target space). Σ stands collectively for non-dynamical entities (background structures) whose values are fixed by means independent of equation (3). It could, for example, be the Minkowski metric in Maxwell's equations and also external currents. The meaning of (3) is to determine Φ, *given* Σ (and the boundary conditions for Φ). We stress that Σ is a constitutive part of the equations of motion. We now make the following

Definition 1. *An action of the group G on the space of dynamical entities Φ is is said to correspond to a* symmetry *of the equations of motion iff*[2] *for all $g \in G$ we have*

$$\mathcal{E}[\Sigma, \Phi] = 0 \iff \mathcal{E}[\Sigma, g \cdot \Phi] = 0 \,. \tag{4}$$

Different form that is mere "covariance", which is a far more trivial requirement. It arises if the space of background structures, Σ, also carries an action of G (as it naturally does if the Σ are tensor fields). Then we have

Definition 2. *An action of the group G on the space of dynamical and non-dynamical entities Φ and Σ is is said to correspond to a* covariance *of the equations of motion, iff for all $g \in G$ we have*

$$\mathcal{E}[\Sigma, \Phi] = 0 \iff \mathcal{E}[g \cdot \Sigma, g \cdot \Phi] = 0 \,. \tag{5}$$

The difference to symmetries being that the background structures – and in that sense the equations of motion themselves – are changed too. Equation (4) says that if Φ solves the equations of motion then $g \cdot \Phi$ solves the *very same* equations. In contrast, (5) merely tells us that if Φ solves the equations of motion, then $g \cdot \Phi$ solves the appropriately transformed equations. Trivially, a symmetry is also a covariance but the converse it not true. Rather, a covariance is a symmetry iff it stabilizes the background structures, i.e. if for all $g \in G$ we have $g \cdot \Sigma = \Sigma$.

Usually one has a good idea of what the dynamical entities Φ in ones theory should be, whereas the choice of Σ is more a matter of presentation and therefore conventional. After all, the only task of equations of motion is to characterize the set $\mathcal{S}$ of dynamically possible fields (and particles) amongst the set $\mathcal{K}$ of all kinematically possible ones. Whether this is done by using auxiliary structures

[2] We write "iff" as abbreviation for "if and only if".

Σ_1 or Σ_2 does not affect the physics. It is for this reason that one has to regard the requirement of mere covariance as, physically speaking, rather empty. This is because one can always achieve covariance by suitably adding non-dynamical structures Σ. Let us give an example for this.

Consider the familiar heat equation,

$$\partial_t T - \kappa \Delta T = 0\,. \tag{6}$$

Here the dynamical field $\Phi = T$ is the temperature function. The background structure is the 3-dimensional Euclidean metric of space which enters the Laplacian, $\Delta := \delta^{ab}\partial_a\partial_b$; hence $\Sigma = \delta$. This equation possesses time translations and Euclidean motions in space (we neglect space reflections for simplicity) as symmetries. These form the group $E_3 \cong \mathbb{R}^3 \rtimes SO(3)$, the semi-direct product of spatial translations and rotations. Clearly E_3 stabilizes δ.

But without changing the physics we can rewrite (6) in the following spacetime form: Let $(x^0, x^1, x^2, x^3) = (ct, x, y, z)$ be inertial coordinates in Minkowski space and $n = \partial_t = n^\mu\partial_\mu$ (i.e. $n^\mu = (c, 0, 0, 0)$) the constant vector field describing the motion of the inertial observer. The components of the Minkowski metric in these coordinates are denoted by $g_{\mu\nu}$. In our conventions ('mostly minus') $\{g_{\mu\nu}\} = \mathrm{diag}(1,\text{-}1,\text{-}1,\text{-}1)$. Then (6) is clearly just the same as

$$n^\mu\partial_\mu T - \kappa\left(c^{-2}\, n^\mu n^\nu - g^{\mu\nu}\right)\,\partial_\mu\partial_\nu\, T = 0\,. \tag{7}$$

Here the dynamical variable is still $\Phi = T$ but the background variables are now given by $\Sigma = (n, g)$. This equations is now manifestly covariant under the Lorentz group if n^μ and $g^{\mu\nu}$ are acted upon as indicated by their indices. Hence we were able to enlarge the covariance group by enlarging the space of Σs. In fact, we could even make the equation covariant under general diffeomorphisms by replacing partial with covariant derivatives. But note that the symmetry group would still be that subgroup that stabilizes (leaves invariant) the (flat) metric g and the (covariant constant) vector field n, which again results in the same symmetry group as for the original equation (6).

3 The Impact of the Relativity Principle on the Automorphism Group of Spacetime

In the history of SR it has often been asked what the most general transformations of spacetime were that implemented the relativity principle (RP), without making use of the requirement of the constancy of the speed of light. This question was first addressed by Ignatowsky [30], who showed that under a certain set of technical assumptions (not consistently spelled out by him) the RP alone suffices to arrive at a spacetime symmetry group which is either the Galilei or the Lorentz group, the latter for some yet undetermined limiting velocity c. More precisely, what is actually shown in this fashion is, as we will see, that the spacetime symmetry group must contain either the proper orthochronous Galilei or

Lorentz group, if the group is required to comprise at least spacetime translations, spatial rotations, and boosts (velocity transformations). What we hence gain is the group-theoretic insight of how these transformations must combine into a common group, given that they form a group at all. We do not learn anything about other transformations, like spacetime reflections or dilations, whose existence we neither required nor ruled out on this theoretical level.

The work of Ignatowsky was put into a logically more coherent form by Franck & Rothe [19] [20], who showed that some of the technical assumptions could be dropped. Further formal simplifications were achieved by Berzi & Gorini [8]. Below we shall basically follow their line of reasoning, except that we do not impose the continuity of the transformations as a requirement, but conclude it from their preservation of the inertial structure plus bijectivity. See also [3] for an alternative discussion on the level of Lie algebras.

The principles of SR are mathematically most concisely expressed in terms of few simple structures put onto spacetime. In SR these structures are absolute in the sense of not being subject to any dynamical change. From a fundamental point of view, it seems rather a matter of convention whether one thinks of these structures as primarily algebraic or geometric. According to the idea advocated by Felix Klein in his "Erlanger Programm" [35], a geometric structure can be characterized by its *automorphism group*[3]. The latter is generally defined by the subgroup of bijections of the set in question which leaves the geometric structure – e.g. thought of as being given in terms of relations – invariant. Conversely, any transformation group (i.e. subgroup of group of bijections) can be considered as the automorphism group of some "geometry" which is defined via the invariant relations.

The geometric structure of spacetime is not a priori given to us. It depends on the physical means on which we agree to measure spatial distances and time durations. These means refer to physical systems, like "rods" and "clocks", which are themselves subject to dynamical laws in spacetime. For example, at a fundamental physical level, the spatial transportation of a rod or a clock from one place to another is certainly a complicated *dynamical* process. It is only due to the special definition of "rod" and "clock" that the result of such a process can be summarized by simple *kinematical* rules. Most importantly, their dynamical behavior must be "stable" in the sense of being essentially independent of their dynamical environment. Hence there is always an implicit consistency hypothesis underlying operational definitions of spatio-temporal measurements, which in case of SR amount to the assumption that rods and clocks are themselves governed by Lorentz invariant dynamical laws.

A basic physical law is the law of inertia. It states the preference of certain types of motions for force-free, uncharged, zero-spin test-particles: the "uniform" and "rectilinear" ones. In the spacetime picture this corresponds to the preference of certain curves corresponding to the inertial worldlines of the force-free test particles. In the gravity-free case, we model these world lines by straight lines of the affine space $\mathrm{Aff}(\mathbb{R}^4)$ over the vector space $\mathbb{R}^4$. This closely corresponds to

[3] Klein calls it "Hauptgruppe".

our intuitive notion of homogeneity of space and time, that is, that there exists an effective and transitive (and hence simply transitive) action of the Abelian group $\mathbb{R}^4$ of translations (cf. Sects. A.1 and A.7). A lot could (and perhaps should) be said at this point about the proper statement of the law of inertia and precisely how it endows spacetimes with certain geometric structures. Instead we will simply refer the interested reader to the literature; see e.g. [25] and references therein.

Note that we do not conversely assume any straight line to correspond to some inertial world-line. Hence the first geometric structure on spacetime, which can be thought of as imposed by the law of inertia, is that of a subset of straight lines. If all straight lines were involved, the automorphism group of spacetime would necessarily have to map any straight line to a straight line and therefore be a subgroup of the affine group $\mathbb{R}^4 \rtimes \mathsf{GL}(4,\mathbb{R})$. This is just the content of the main theorem in affine geometry; see e.g. [6]. However, we can only argue that it must map the subset of inertial world-lines onto itself. We take this subset to consist of all straight lines in $\mathrm{Aff}(\mathbb{R}^4)$ whose slope with respect to some reference direction is smaller than a certain finite value β. This corresponds to all worldlines not exceeding a certain limiting speed with reference to some inertial frame. It is then still true that any bijection[4] of $\mathrm{Aff}(\mathbb{R}^4)$ preserving that subset must be a subgroup of the affine group [28]. Also, it is not necessary to assume that lines map surjectively onto lines [16].

For further determination of the automorphism group of spacetime we invoke the following principles:

ST1: Homogeneity of spacetime.
ST2: Isotropy of space.
ST3: Galilean principle of relativity.

We take ST1 to mean that the sought-for group should include all translations and hence be of the form $\mathbb{R}^4 \rtimes \mathsf{G}$, where G is a subgroup of $\mathsf{GL}(4,\mathbb{R})$. ST2 is interpreted as saying that G should include the set of all spatial rotations. If, with respect to some frame, we write the general element $A \in \mathsf{GL}(4,\mathbb{R})$ in a $1+3$ split form (thinking of the first coordinate as time, the other three as space), we want G to include all

$$R(\mathbf{D}) = \begin{pmatrix} 1 & \mathbf{0}^\top \\ \mathbf{0} & \mathbf{D} \end{pmatrix}, \qquad \text{where} \quad \mathbf{D} \in \mathsf{SO}(3)\,. \tag{8}$$

[4] If one drops the assumption of bijectivity, then there exist in addition the fractional linear transformations which map straight lines to straight lines, except for those points that are mapped to 'infinity'; see e.g. the discussion in Fock's book [18], in particular his Appendix A, and also [20]. One might argue that since physics takes place in the finite we cannot sensibly argue for global bijectivity and hence have to consider those more general transformations. However, the group they generate does not have an invariant bounded domain in spacetime and hence cannot be considered as the automorphism group of any fixed set of physical events.

Finally, ST3 says that velocity transformations, henceforth called "boosts", are also contained in G. However, at this stage we do not know how boosts are to be represented mathematically. Let us make the following assumptions:

B1: Boosts $B(\mathbf{v})$ are labeled by a vector $\mathbf{v} \in B_c(\mathbb{R}^3)$, where $B_c(\mathbb{R}^3)$ is the open ball in $\mathbb{R}^3$ of radius c. The physical interpretation of $\mathbf{v}$ shall be that of the boost velocity, as measured in the system from which the transformation is carried out. We allow c to be finite or infinite ($B_\infty(\mathbb{R}^3) = \mathbb{R}^3$). $\mathbf{v} = \mathbf{0}$ corresponds to the identity transformation, i.e. $B(\mathbf{0}) = \mathrm{id}_{\mathbb{R}^4}$. We also assume that $\mathbf{v}$, considered as coordinate function on the group, is continuous.

B2: As part of ST2 we require equivariance of boosts under rotations:

$$R(\mathbf{D}) \cdot B(\mathbf{v}) \cdot R(\mathbf{D}^{-1}) = B(\mathbf{D} \cdot \mathbf{v})\,. \tag{9}$$

The latter assumption allows us to restrict attention to boost in a fixed direction, say that of the positive x-axis. Once their analytical form is determined as function of v, where $\mathbf{v} = v\mathbf{e}_x$, we deduce the general expression for boosts using (9) and (8). We make no assumptions involving space reflections.[5] We now restrict attention to $\mathbf{v} = v\mathbf{e}_x$. We wish to determine the most general form of $B(\mathbf{v})$ compatible with all requirements put so far. We proceed in several steps:

1. Using an arbitrary rotation $\mathbf{D}$ around the x-axis, so that $\mathbf{D} \cdot \mathbf{v} = \mathbf{v}$, equation (9) allows to prove that

$$B(v\mathbf{e}_x) = \begin{pmatrix} \mathbf{A}(v) & 0 \\ 0 & \alpha(v)\mathbf{1}_2 \end{pmatrix}\,, \tag{10}$$

   where here we wrote the 4×4 matrix in a $2+2$ decomposed form. (i.e. $\mathbf{A}(v)$ is a 2×2 matrix and $\mathbf{1}_2$ is the 2×2 unit-matrix). Applying (9) once more, this time using a π-rotation about the y-axis, we learn that α is an even function, i.e.

$$\alpha(v) = \alpha(-v)\,. \tag{11}$$

   Below we will see that $\alpha(v) \equiv 1$.

2. Let us now focus on $\mathbf{A}(v)$, which defines the action of the boost in the $t - x$ plane. We write

$$\begin{pmatrix} t \\ x \end{pmatrix} \mapsto \begin{pmatrix} t' \\ x' \end{pmatrix} = \mathbf{A}(v) \cdot \begin{pmatrix} t \\ x \end{pmatrix} = \begin{pmatrix} a(v) & b(v) \\ c(v) & d(v) \end{pmatrix} \cdot \begin{pmatrix} t \\ x \end{pmatrix}\,. \tag{12}$$

   We refer to the system with coordinates (t, x) as K and that with coordinates (t', x') as K'. From (12) and the inverse (which is elementary to compute)

[5] Some derivations in the literature of the Lorentz group do not state the equivariance property (9) explicitly, though they all use it (implicitly), usually in statements to the effect that it is sufficient to consider boosts in one fixed direction. Once this restriction is effected, a one-dimensional spatial reflection transformation is considered to relate a boost transformation to that with opposite velocity. This then gives the impression that reflection equivariance is also invoked, though this is not necessary, for (9) allows to invert one axis through a 180-degree rotation about a perpendicular one.

one infers that the velocity v of K' with respect to K and the velocity v' of K with respect to K' are given by

$$v = -c(v)/d(v)\,, \tag{13a}$$

$$v' = -v\,d(v)/a(v) =: \varphi(v)\,. \tag{13b}$$

Since the transformation $K' \to K$ is the inverse of $K \to K'$, the function $\varphi : (-c, c) \to (-c, c)$ obeys

$$\mathbf{A}(\varphi(v)) = (\mathbf{A}(v))^{-1}\,. \tag{14}$$

Hence φ is a bijection of the open interval $(-c, c)$ onto itself and obeys

$$\varphi \circ \varphi = \mathrm{id}_{(-c,c)}\,. \tag{15}$$

3. Next we determine φ. Once more using (9), where $\mathbf{D}$ is a π-rotation about the y-axis, shows that the functions a and d in (10) are even and the functions b and c are odd. The definition (13b) of φ then implies that φ is odd. Since we assumed $\mathbf{v}$ to be a continuous coordinatization of a topological group, the map φ must also be continuous (since the inversion map, $g \mapsto g^{-1}$, is continuous in a topological group). A standard theorem now states that a continuous bijection of an interval of $\mathbb{R}$ onto itself must be strictly monotonic. Together with (15) this implies that φ is either the identity or minus the identity map. If it is the identity map, evaluation of (14) shows that either the determinant of $\mathbf{A}(v)$ must equals -1, or that $\mathbf{A}(v)$ is the identity for all $\mathbf{v}$. We exclude the second possibility straightaway and the first one on the grounds that we required $\mathbf{A}(v)$ be the identity for $v = 0$. Also, in that case, (14) implies $A^2(v) = \mathrm{id}$ for all $v \in (-c, c)$. We conclude that $\varphi = -\mathrm{id}$, which implies that the relative velocity of K with respect to K' is minus the relative velocity of K' with respect to K. Plausible as it might seem, there is no a priori reason why this should be so.[6] The RP only implies (15), not the stronger relation $\varphi(v) = -v$. This was first pointed out in [8]
4. We briefly revisit (11). Since we have seen that $B(-v\mathbf{e}_x)$ is the inverse of $B(v\mathbf{e}_x)$, we must have $\alpha(-v) = 1/\alpha(v)$, so that (11) implies $\alpha(v) \equiv \pm 1$. But only $\alpha(v) \equiv +1$ is compatible with our requirement that $B(\mathbf{0})$ be the identity.
5. Now we return to the determination of $\mathbf{A}(v)$. Using (13) and $\varphi = -\mathrm{id}$, we write

$$\mathbf{A}(v) = \begin{pmatrix} a(v) & b(v) \\ -va(v) & a(v) \end{pmatrix} \tag{16}$$

and

$$\Delta(v) := \det\big(\mathbf{A}(v)\big) = a(v)\big[a(v) + vb(v)\big]\,. \tag{17}$$

Equation $\mathbf{A}(-v) = (\mathbf{A}(v))^{-1}$ is now equivalent to

[6] Note that v and v' are measured with different sets of rods and clocks.

$$a(-v) = a(v)/\Delta(v)\,, \tag{18a}$$
$$b(-v) = -\,b(v)/\Delta(v)\,. \tag{18b}$$

Since, as already seen, a is an even and b is an odd function, (18) is equivalent to $\Delta(v) \equiv 1$, i.e. the unimodularity of $B(\mathbf{v})$. Equation (17) then allows to express b in terms of a:

$$b(v) = \frac{a(v)}{v}\left[\frac{1}{a^2(v)} - 1\right]\,. \tag{19}$$

6. Our problem is now reduced to the determination of the single function a. This we achieve by employing the requirement that the composition of two boosts in the same direction results again in a boost in that direction, i.e.

$$\mathbf{A}(v)\cdot\mathbf{A}(v') = \mathbf{A}(v'')\,. \tag{20}$$

According to (16) each matrix $\mathbf{A}(v)$ has equal diagonal entries. Applied to the product matrix on the left hand side of (20) this implies that $v^{-2}(a^{-2}(v) - 1)$ is independent of v, i.e. equal to some constant k whose physical dimension is that of an inverse velocity squared. Hence we have

$$a(v) = \frac{1}{\sqrt{1 + kv^2}}\,, \tag{21}$$

where we have chosen the positive square root since we require $a(0) = 1$. The other implications of (20) are

$$a(v)a(v')(1 - kvv') = a(v'')\,, \tag{22a}$$
$$a(v)a(v')(1 + vv') = v''a(v'')\,, \tag{22b}$$

from which we deduce

$$v'' = \frac{v + v'}{1 - kvv'}\,. \tag{23}$$

Conversely, (21) and (23) imply (22). We conclude that (20) is equivalent to (21) and (23).

7. So far a boost in x direction has been shown to act non-trivially only in the $t-x$ plane, where its action is given by the matrix that results from inserting (19) and (21) into (16):

$$\mathbf{A}(v) = \begin{pmatrix} a(v) & kv\,a(v) \\ -v\,a(v) & a(v)\,, \end{pmatrix} \qquad \text{where} \quad a(v) = 1/\sqrt{1 + kv^2}\,. \tag{24}$$

- If $k > 0$ we rescale $t \mapsto \tau := t/\sqrt{k}$ and set $\sqrt{k}\,v := \tan\alpha$. Then (24) is seen to be a Euclidean rotation with angle α in the $\tau - x$ plane. The velocity spectrum is the whole real line plus infinity, i.e. a circle, corresponding to $\alpha \in [0, 2\pi]$, where 0 and 2π are identified. Accordingly, the composition law (23) is just ordinary addition for the angle α. This causes several paradoxa when v is interpreted as velocity. For example,

composing two finite velocities v, v' which satisfy $vv' = 1/k$ results in $v'' = \infty$, and composing two finite and positive velocities, each of which is greater than $1/\sqrt{k}$, results in a finite but negative velocity. In this way the successive composition of finite positive velocities could also result in zero velocity. The group $\mathsf{G} \subset \mathsf{GL}(n, \mathbb{R})$ obtained in this fashion is, in fact, $\mathsf{SO}(4)$. This group may be uniquely characterized as the largest connected group of bijections of $\mathbb{R}^4$ that preserves the Euclidean distance measure. In particular, it treats time symmetrically with all space directions, so that no invariant notion of time-orientability can be given in this case.

- For $k = 0$ the transformations are just the ordinary boosts of the Galilei group. The velocity spectrum is the whole real line (i.e. v is unbounded but finite) and G is the Galilei group. The law for composing velocities is just ordinary vector addition.
- Finally, for $k < 0$, one infers from (23) that $c := 1/\sqrt{-k}$ is an upper bound for all velocities, in the sense that composing two velocities taken from the interval $(-c, c)$ always results in a velocity from within that interval. Writing $\tau := ct$, $v/c =: \beta =: \tanh\rho$, and $\gamma = 1/\sqrt{1-\beta^2}$, the matrix (24) is seen to be a *Lorentz boost* or *hyperbolic motion* in the $\tau - x$ plane:

$$\begin{pmatrix} \tau \\ x \end{pmatrix} \mapsto \begin{pmatrix} \gamma & -\beta\gamma \\ -\beta\gamma & \gamma \end{pmatrix} \cdot \begin{pmatrix} \tau \\ x \end{pmatrix} = \begin{pmatrix} \cosh\rho & -\sinh\rho \\ -\sinh\rho & \cosh\rho \end{pmatrix} \cdot \begin{pmatrix} \tau \\ x \end{pmatrix} . \tag{25}$$

The quantity

$$\rho := \tanh^{-1}(v/c) = \tanh^{-1}(\beta) \tag{26}$$

is called *rapidity*.[7] If rewritten in terms of the corresponding rapidities the composition law (23) reduces to ordinary addition: $\rho'' = \rho + \rho'$.

This shows that only the Galilei and the Lorentz group survive as candidates for any symmetry group implementing the RP. Once the Lorentz group for velocity parameter c is chosen, one may prove that it is fully characterized by its property to leave a certain symmetric bilinear form invariant (cf. Sect A.4). Endowing spacetime with that structure plus the affine structure from the law of inertia, we can characterize the Lorentz group as automorphism group of some geometric structure. This is often the starting point of more axiomatic approaches. Here we preferred to start with the opposite strategy, which stresses that the geometry of spacetime is a contingent physical property, emerging through its automorphism group, which in turn relates to the actual dynamical laws of nature. Having said that, we may now follow the convenient axiomatic line of presentation.

4 Algebraic Structures of Minkowski Space

Definition 3. Minkowski space *is the affine space* $\mathrm{Aff}(\mathbb{R}^4)$ *over the four-dimensional real vector space* $\mathbb{R}^4$*, where the latter is endowed with a symmetric*

[7] This term was coined by Robb [39], but the quantity was used before by others; compare [50].

non-degenerate bilinear form g of signature $(+,-,-,-) = (1,3)$. We write $\mathbb{M}^4 = (\mathrm{Aff}(\mathbb{R}^4), g)$. We shall usually restrict to bases $\{e_\mu\}_{\mu=0\cdots 3}$ of $\mathbb{R}^4$ for which $g(e_\mu, e_\nu) =: g_{\mu\nu} = diag(1,-1,-1,-1)$.

4.1 The Lorentz and the Galilei Group

Definition 4. *The* (homogeneous) Lorentz group *is the linear group (subgroup of $\mathsf{GL}(4,\mathbb{R})$) of orthogonal transformations of $\mathbb{M}^4$, also called $\mathsf{O}(1,3)$. Hence $\{L^\mu_\nu\} \in \mathsf{O}(1,3)$ iff*

$$g_{\mu\nu} L^\mu_\alpha L^\nu_\beta = g_{\alpha\beta}\,. \tag{27}$$

Note that according to Proposition 9 orthogonal transformations are necessarily linear.

As topological space $\mathsf{O}(1,3)$ decomposes into the disjoint union of four connected components. Here $+/-$ stands for positive/negative determinant and $\uparrow$ / $\downarrow$ for time-orientation preserving/reversing respectively:

$$\mathsf{O}(1,3) = \underbrace{\mathsf{O}^\uparrow_+(1,3) \;\cup\; \mathsf{O}^\downarrow_+(1,3)}_{\mathsf{SO}(1,3)} \cup \mathsf{O}^\uparrow_-(1,3) \;\cup\; \mathsf{O}^\downarrow_-(1,3)\,. \tag{28}$$

Of these four components only $\mathsf{O}^\uparrow_+(1,3)$, the component containing the group identity, is a subgroup, called the group of proper orthochronous Lorentz transformations. Elementwise composition with space/time reflections gives $\mathsf{O}^\uparrow_-(1,3)/\mathsf{O}^\downarrow_-(1,3)$ respectively. In the sequel we shall also write Lor for $\mathsf{O}(1,3)$ and $\mathsf{Lor}^\uparrow_+$ for $\mathsf{O}^\uparrow_+(1,3)$.

For any group $\mathsf{G} \subset \mathsf{GL}(n,\mathbb{R})$, there is a corresponding *inhomogeneous group*, IG, given by the semi-direct product

$$\mathsf{IG} = \{(a,A) \mid a \in \mathbb{R}^n\,,\; A \in \mathsf{G}\}\,, \tag{29}$$

where

$$(a,A)(a',A') = (a + A\cdot a'\,,\; A\cdot A')\,. \tag{30}$$

It can again be thought of as subgroup of $\mathsf{GL}(n+1\,,\,\mathbb{R})$ via the embedding

$$(a,A) \longmapsto \begin{pmatrix} 1 & 0^\top \\ a & A \end{pmatrix}\,. \tag{31}$$

In this fashion we get the inhomogeneous Lorentz groups ILor and $\mathsf{ILor}^\uparrow_+$ also called Poincaré groups.

Let us recall the structure of the proper orthochronous homogeneous Galilei group, which we denote by $\mathsf{Gal}^\uparrow_+$. It is generated by spatial rotations $\mathbf{x} \mapsto \mathbf{x}' = \mathbf{D}\cdot\mathbf{x}$ and boosts $\mathbf{x} \mapsto \mathbf{x}' = \mathbf{x} + \mathbf{v}t$ ($t' = t$ in both cases). Hence, if we agree to let rotations act first and then act with the boosts, the general form of a matrix in $\mathsf{Gal}^\uparrow_+ \subset \mathsf{GL}(4,\mathbb{R})$ will be (written in a $1+3$ decomposition):

$$G(\mathbf{v},\mathbf{D})) := \begin{pmatrix} 1 & \mathbf{0}^\top \\ \mathbf{v} & \mathbf{1}_3 \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0}^\top \\ \mathbf{0} & \mathbf{D} \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{0}^\top \\ \mathbf{v} & \mathbf{D} \end{pmatrix}. \tag{32}$$

Hence, given any pair $(\mathbf{v},\mathbf{D})$, this tells us how to uniquely construct the matrix $G(\mathbf{v},\mathbf{D}) \in \mathsf{Gal}^\uparrow_+$. Conversely, given a matrix $G \in \mathsf{Gal}^\uparrow_+$, we can immediately tell $\mathbf{v} \in \mathbb{R}^3$ and $\mathbf{D} \in \mathsf{SO}(3)$ by comparison with the general form (32). Hence there is a bijection of sets $G : \mathbb{R}^3 \times \mathsf{SO}(3) \to \mathsf{Gal}^\uparrow_+$. The group structure on $\mathbb{R}^3 \times \mathsf{SO}$ that makes this into an isomorphism of groups is a semi-direct product:

$$G(\mathbf{v}_1,\mathbf{D}_1) \cdot G(\mathbf{v}_2,\mathbf{D}_2) = G(\mathbf{v}_1 + \mathbf{D}_1 \cdot \mathbf{v}_2\,,\, \mathbf{D}_1 \cdot \mathbf{D}_2)\,. \tag{33}$$

Hence we have an isomorphism $\mathsf{Gal}^\uparrow_+ \cong \mathbb{R}^3 \rtimes \mathsf{SO}(3)$. This also follows straightaway from comparing (31) with (32). From (33) the law for taking the inverse is easily deduced:

$$\big(G(\mathbf{v},\mathbf{D})\big)^{-1} = G(-\mathbf{D}^{-1} \cdot v\,,\, \mathbf{D}^{-1}) \tag{34}$$

The inhomogeneous Galilei group is now isomorphic to an iterated semi direct product:

$$\mathsf{IGal}^\uparrow_+ := \mathbb{R}^4 \rtimes \mathsf{Gal}^\uparrow_+ \cong \mathbb{R}^4 \rtimes (\mathbb{R}^3 \rtimes \mathsf{SO}(3))\,, \tag{35}$$

where $\mathbb{R}^4$ corresponds to space-time translations and $\mathbb{R}^3$ to boost. The action of $\mathsf{Gal}^\uparrow_+$ on $\mathbb{R}^4$ is via the "defining representation", i.e. the obvious action of 4×4 matrices of the form (32) on $\mathbb{R}^4$. Not that this 4-dimensional representation of $\mathsf{Gal}^\uparrow_+$ is reducible: it transforms the 3-dimensional subspace of "spatial" vectors $(0,\mathbf{a})^\top$ into themselves. Hence the semi-direct product of $\mathsf{Gal}^\uparrow_+$ with the subgroup of pure spatial translations, isomorphic to $\mathbb{R}^3$, is a proper subgroup of $\mathsf{IGal}^\uparrow_+$ that properly contains $\mathsf{Gal}^\uparrow_+$: $\mathsf{Gal}^\uparrow_+ \subset \mathbb{R}^3 \rtimes \mathsf{Gal}^\uparrow_+ \subset \mathsf{IGal}^\uparrow_+$. In other words: $\mathsf{Gal}^\uparrow_+$ is *not* a maximal[8] subgroup of $\mathsf{IGal}^\uparrow_+$. Hence another way to write $\mathsf{IGal}^\uparrow_+$ as semi-direct product is

$$\mathsf{IGal}^\uparrow_+ \cong (\mathbb{R}^3 \times \mathbb{R}^3) \rtimes (\mathbb{R} \times \mathsf{SO}(3))\,. \tag{36}$$

where the first two $\mathbb{R}^3$ on the right hand side correspond to spatial translations and boosts respectively, and the single $\mathbb{R}$ to time translations. The action of $\mathbb{R} \times \mathsf{SO}(3)$ on $\mathbb{R}^3 \times \mathbb{R}^3$ is the factor-wise standard action of $\mathsf{SO}(3)$ on $\mathbb{R}^3$ and the trivial action of $\mathbb{R}$.

At this point we can already anticipate some major group-theoretic differences between the Galilei and the Lorentz groups (denoted by Lor). For example:

1. $\mathsf{Lor}^\uparrow_+$ is a simple group, that is, it does not contain any normal subgroup other than the trivial ones (itself and the unit element). The set of pure boost does not form a subgroup. In contrast, $\mathsf{Gal}^\uparrow_+$ is not even semi-simple, meaning that it contains a non-trivial Abelian normal subgroup, namely the boosts. This makes a big difference in the corresponding representation theories.

[8] A proper subgroup $G' \subsetneq G$ is *maximal* if there is no subgroup H of G such that $G' \subsetneq H \subsetneq G$.

2. In $\mathsf{ILor}^{\uparrow}_{+} = \mathbb{R}^4 \rtimes \mathsf{Lor}^{\uparrow}_{+}$ the action of $\mathsf{Lor}^{\uparrow}_{+}$ on $\mathbb{R}^4$ is irreducible and $\mathsf{Lor}^{\uparrow}_{+}$ is a maximal subgroup of $\mathsf{ILor}^{\uparrow}_{+}$, in contrast to the Galilean case. This makes a difference for the existence of invariant equivalence relations on spacetime (cf. Sect. A.1), like, for example, absolute simultaneity structures. This will be further discussed in Sect. 5.3.

4.2 Polar Decomposition

In (32) we have given an easy proof-by-inspection of the unique decomposability of any element in $\mathsf{Gal}^{\uparrow}_{+}$ into a product of a rotation and a boost. We now like to discuss the analog of this decomposition within $\mathsf{Lor}^{\uparrow}_{+}$, which is more difficult to obtain. We start by recalling the statement and proof of the 'polar decomposition' of matrices:

Proposition 1. *Let $X \in \mathsf{GL}(n,\mathbb{C})$; then there exists a unique $R \in \mathsf{U}(n)$ (i.e. $R^\dagger = R^{-1}$) and a unique positive-definite Hermitian matrix B (i.e. $B = B^\dagger$ with strictly positive eigenvalues) such that*

$$X = B \cdot R\,. \tag{37}$$

If $X \in \mathsf{GL}(n,\mathbb{R})$ then B is real, symmetric, and positive definite. R is real and orthogonal.

Proof. Let $A := XX^\dagger$, which is positive-definite and Hermitean (zero eigenvalues are excluded since X is invertible). Recall that the square-root is a well defined bijective map (a homeomorphism in fact) of the space of positive-definite Hermitean matrices onto itself. Define $B := \sqrt{A}$ and $R := B^{-1}X$, then $R^\dagger = X^\dagger B^{-1} = X^{-1}B = R^{-1}$, where the first equality follows from Hermiticity of B and the second from $B^2 = XX^\dagger$. Hence R is unitary and we have shown existence of a polar decomposition. To show uniqueness, assume there exist two such decompositions: $X = B_1R_1 = B_2R_2$. Then $B_1 = B_2R_3$, where $R_3 := R_2R_1^{-1}$ is again unitary. Hermiticity of $B_{1,2}$ and unitarity of R_3 now imply $B_1^2 = B_1B_1^\dagger = B_2R_3R_3^\dagger B_2^\dagger = B_2^2$ and hence $B_1 = B_2$, since "squaring" is an injective map (a homeomorphism in fact) from the space of positive-definite Hermitean matrices onto itself. This, in turn, implies $R_1 = R_2$ and hence uniqueness. Finally, if X is real, then B and consequently R are also real. □

We wish to apply this to $\mathsf{Lor}^{\uparrow}_{+} \subset \mathsf{GL}(4,\mathbb{R})$. But note that polar decomposing an element in $\mathsf{G} \subset \mathsf{GL}(n,\mathbb{C})$ need not generally lead to factors in G. However, this is true in many cases. For example, we have

Proposition 2. *Let $E^{(p,q)}$ be the diagonal matrix whose first p diagonal entries equal $+1$ and the remaining $q = n - p$ diagonal entries equal -1. We define the group*

$$\mathsf{U}(p,q) := \{X \in \mathsf{GL}(n,\mathbb{C}) \;\mid\; X \cdot E^{(p,q)} \cdot X^\dagger = E^{(p,q)}\}\,. \tag{38}$$

Restricting to matrices with real entries gives the group $\mathsf{O}(p,q)$. Polar decomposing elements of $\mathsf{U}(p,q)$ or $\mathsf{O}(p,q)$ leads to factors within these groups respectively. The same is true if we restrict to the identity components of these groups.

Proof. It is sufficient to prove that $B = \sqrt{XX^\dagger}$ is in $\mathsf{U}(p,q)$ since this clearly implies that the product $R = B^{-1}X$ will also be in $\mathsf{U}(p,q)$. Now, $E^{(p,q)} \in \mathsf{U}(p,q)$ (it clearly satisfies the defining relation in (38)) so that $X^\dagger$ and hence $XX^\dagger$ are elements in $\mathsf{U}(p,q)$. But then $\sqrt{XX^\dagger} \in \mathsf{U}(p,q)$, too. To see this, use e.g. the exponential map (cf. Sect. A.10), which defines a homeomorphism from the space of Hermitean to the space of positive-definite Hermitean matrices. Then $X = \exp(Y) \in \mathsf{U}(p,q) \Leftrightarrow E^{(p,q)} \cdot Y \cdot E^{(p,q)} = -Y^\dagger \Leftrightarrow \sqrt{X} = \exp(Y/2) \in \mathsf{U}(p,q)$. Finally it is clear from the explicit construction of the polar factors that if $X(s)$ is a continuous path connecting the identity to X, and if $X(s) = B(s)R(s)$ is the polar decomposition for each value of s, then $B(s)$ and $R(s)$ are continuous paths connecting B and R to the identity. □

We will use this to decompose any proper orthochronous Lorentz transformation L into a boost B and a proper spatial rotation R.[9] Let

$$L = \begin{pmatrix} \gamma & \mathbf{a}^\top \\ \mathbf{b} & \mathbf{M} \end{pmatrix} \tag{39}$$

be a Lorentz transformation. The defining relation (27), as well as the relation $L^\alpha_\mu L^\beta_\nu g^{\mu\nu} = g^{\alpha\beta}$ which follows from it, are equivalent respectively to

$$\mathbf{a}^2 = \gamma^2 - 1\,, \quad \gamma\mathbf{b} = \mathbf{M}\cdot\mathbf{a}\,, \qquad \mathbf{M}\cdot\mathbf{M}^\top = \mathbf{1}_3 + \mathbf{b}\otimes\mathbf{b}^\top\,, \tag{40a}$$

$$\mathbf{b}^2 = \gamma^2 - 1\,, \quad \gamma\mathbf{a} = \mathbf{M}^\top\cdot\mathbf{b}\,, \quad \mathbf{M}^\top\cdot\mathbf{M} = \mathbf{1}_3 + \mathbf{a}\otimes\mathbf{a}^\top\,. \tag{40b}$$

The polar decomposition of the matrix $L \in \mathsf{O}(1,3)^\uparrow_+$ in (39) is given by

$$L = B\cdot R \tag{41}$$

with

$$B = \begin{pmatrix} \gamma & \mathbf{b}^\top \\ \mathbf{b} & \mathbf{1}_3 + \frac{\mathbf{b}\otimes\mathbf{b}^\top}{1+\gamma} \end{pmatrix}\,, \qquad R = \begin{pmatrix} 1 & \mathbf{0}^\top \\ \mathbf{0} & \mathbf{M} - \frac{\mathbf{b}\otimes\mathbf{a}^\top}{1+\gamma} \end{pmatrix}\,. \tag{42}$$

Tho check this, first verify that L is indeed the product $B\cdot R$ using the relations (40). Next we note that B is symmetric and that its eigenvalues (EV) are all positive:

$$\mathrm{EV}(B) = (\gamma + \sqrt{\gamma^2-1}\,,\, \gamma - \sqrt{\gamma^2-1}\,,\, 1\,,\, 1) \;>\; 0\,. \tag{43}$$

Finally one checks that R is a spatial rotation, i.e.

$$\mathbf{D} := \mathbf{M} - \frac{\mathbf{b}\otimes\mathbf{a}^\top}{1+\gamma} \;\in\; \mathsf{SO}(3)\,. \tag{44}$$

[9] Note that the analogous factorization (32) of a homogeneous Galilei transformation into boost and rotation is *not* given by polar decomposition, but rather by a decomposition into a lower triangular matrix with unit diagonal (the boost) and an orthogonal matrix. This is a special case of what is generally known as Iwasawa decomposition.

Indeed, $\mathbf{D}\cdot\mathbf{D}^\top = \mathbf{1}_3$ is easily verified using the relations (40) and $\det(\mathbf{D}) = 1$ follows from $\det(L) = \det(B) = 1$. Hence we have found the polar decomposition of $L \in \mathsf{O}^\uparrow_+(1,3)$.

We can now characterize the factors B (boost) and R (rotation) of L in terms of the parameters $\gamma, \mathbf{a}, \mathbf{b}, \mathbf{M}$ in (39). We start with R: Using the first and second equation in (40a) one readily shows that

$$\mathbf{D}\cdot\mathbf{a} = \mathbf{b}\,. \tag{45}$$

Hence the plane of rotation for $\mathbf{D}$ is $\mathrm{span}\{\mathbf{a},\mathbf{b}\} \subset \mathbb{R}^3$. The rotation angle θ obeys

$$\cos\theta = \frac{\mathbf{a}\cdot\mathbf{b}}{\gamma^2-1}\,, \tag{46a}$$

where we used $\mathbf{a}^2 = \mathbf{b}^2 = \gamma^2 - 1$ (first equations in (40)). On the other hand, it evidently also obeys the general equation $1 + 2\cos\theta = \mathrm{trace}(\mathbf{D})$, i.e.

$$1 + 2\cos\theta = \mathrm{trace}(\mathbf{M}) - \frac{\mathbf{a}\cdot\mathbf{b}}{1+\gamma}\,. \tag{46b}$$

Elimination of $\mathbf{a}\cdot\mathbf{b}$ via (46a) gives

$$\cos\theta = \frac{\mathrm{trace}(\mathbf{M})-1}{1+\gamma}\,. \tag{46c}$$

Next we set $\boldsymbol{\beta} := \mathbf{b}/\gamma$, $\beta := \|\boldsymbol{\beta}\|$, and $\hat{\boldsymbol{\beta}} := \boldsymbol{\beta}/\beta$; then

$$\gamma = \gamma(\boldsymbol{\beta}) := 1/\sqrt{1-\beta^2}\,, \qquad \mathbf{b} = \gamma\boldsymbol{\beta}\,, \qquad \mathbf{a} = \gamma\mathbf{D}^\top\cdot\boldsymbol{\beta}\,. \tag{47}$$

Writing B in terms of $\boldsymbol{\beta}$ explicitly shows that it is a boost with parameter $\boldsymbol{\beta} = \mathbf{v}/c$.

The general Lorentz transformation (39), instead of being considered as function of $\gamma, \mathbf{a}, \mathbf{b}, \mathbf{M}$ obeying (40), can now be considered as function of $\boldsymbol{\beta}$ and $\mathbf{D}$,

$$L(\boldsymbol{\beta},\mathbf{D}) = \underbrace{\begin{pmatrix} \gamma & \gamma\boldsymbol{\beta}^\top \\ \gamma\boldsymbol{\beta} & \mathbf{1}_3 + (\gamma-1)\,\hat{\boldsymbol{\beta}}\otimes\hat{\boldsymbol{\beta}}^\top \end{pmatrix}}_{=:\,B(\boldsymbol{\beta})} \underbrace{\begin{pmatrix} 1 & \mathbf{0}^\top \\ \mathbf{0} & \mathbf{D} \end{pmatrix}}_{=:\,R(\mathbf{D})}\,, \tag{48}$$

where γ is now understood as function of $\boldsymbol{\beta}$ as in (47). The only restrictions on the parameters being that $\mathbf{D} \in \mathsf{SO}(3)$ and $\boldsymbol{\beta} \in \mathcal{B}_1 \subset \mathbb{R}^3$, where $\mathcal{B}_1$ denotes the ball of unit radius centered at the origin (cf. (69)). The decomposition (48) should be regarded as the analog of (32).

It is easy to check directly that the boost are indeed equivariant with respect to rotations:

$$R(\mathbf{D})\cdot B(\boldsymbol{\beta})\cdot R(\mathbf{D}^{-1}) = B(\mathbf{D}\cdot\boldsymbol{\beta})\,. \tag{49}$$

The polar decomposition is unique once the order of rotations and boosts are fixed. In (41) we had put the rotations to the right, i.e. one first rotates and then boosts (we think actively). Had we chosen the opposite order the rotation parameter would still be $\mathbf{D}$ but the boost parameter would change to $\mathbf{D}^\top\cdot\boldsymbol{\beta}$. This follows immediately from (49).

4.3 The Lie Algebras of the Lorentz and Galilei Groups

The commutation relations of the Lorentz Lie-algebra follow from the general formula (191), where we have to set $\epsilon = 1$. Here we shall rename the generators M_{ab}, where $a, b \in \{0, 1, 2, 3\}$, in the way explained below (indices i, j, k are in $\{1, 2, 3\}$). For direct comparison with (191) we also give their expression in terms of the defining representation (i.e. as elements of $\mathrm{End}(\mathbb{R}^4)$). So let $\{e_a\}_{a=0\cdots3}$ and $\{\eta^a\}_{a=0\cdots3}$ be dual bases of $\mathbb{R}^4$ and $\eta_a := g_{ab}\eta^b$, where $g_{ab} := g(e_a, e_b)$ (cf. Sect. A.5). Then:

$$
\begin{aligned}
J_i &:= \tfrac{1}{2}\varepsilon_{ijk}M_{jk} &&= \varepsilon_{ijk}e_j \otimes \eta_k\,, && \text{(50a)}\\
K_i &:= \tfrac{1}{c}\,M_{i0} &&= \tfrac{1}{c}\,(e_i \otimes \eta_0 - e_0 \otimes \eta_i)\,, && \text{(50b)}\\
P_i &:= T_i &&= e_i\,, && \text{(50c)}\\
E &:= c\,T_0 &&= c\,e_0\,. && \text{(50d)}
\end{aligned}
$$

These generate active rotations, boosts, translations in space, and translations in time respectively. The reason for the factors of $1/c$ in (50b) and c in (50d) is as follows: We wish the K_i to be the generators of boosts with velocity parameters v^i (rather than $\beta^i = v^i/c$), i.e. $\exp(M_{i0}\beta^i) = \exp(K_i v^i)$. Similarly, we wish E to be the generator of time translation with parameter Δt (rather than $\Delta x^0 = c\Delta t$), i.e. $\exp(T_0 \Delta x^0) = \exp(E\Delta t)$. This puts the K_i and E in quantitative analogy to the corresponding generators in the Galilei group and hence facilitates a direct comparison.

The relations (191) now amount to

$$
\begin{aligned}
[J_i, J_j] &= \varepsilon_{ijk}\,J_k && \text{(51a)}\\
[J_i, K_j] &= \varepsilon_{ijk}\,K_k && \text{(51b)}\\
[K_i, K_j] &= -\,\varepsilon_{ijk}\,J_k/c^2 && \text{(51c)}\\
[J_i, P_j] &= \varepsilon_{ijk}\,P_k && \text{(51d)}\\
[J_i, E] &= 0 && \text{(51e)}\\
[K_i, P_j] &= \delta_{ij}\,E/c^2 && \text{(51f)}\\
[K_i, E] &= P_i && \text{(51g)}\\
[P_i, P_j] &= 0 && \text{(51h)}\\
[P_i, E] &= 0\,. && \text{(51i)}
\end{aligned}
$$

Those involving J_i on the left hand side just tell us that the other quantity in the bracket is either a spatial vector or scalar. According to (51a) the J_i form a Lie subalgebra but, as e.g. (51b) shows, not an ideal (cf. Sect. A.9). In contrast, (51c) shows that the K_i do not form a Lie subalgebra. The J_i, K_i span the Lie algebra of $\mathsf{O}(1,3)$ and it is easy to prove from the first three relations above that it is simple (has no non-trivial ideals). Moreover, any of the ten generators appears on the right hand side of some relation (51), i.e. can be written as a commutator. This means that the Lie algebra of the inhomogeneous Lorentz group is *perfect* (i.e. generated by commutators).

Another fact easily seen from (51a-51c) is that the Lie algebra of the homogeneous Lorentz group is the 'complex double' (my terminology, see below) of the Lie algebra of $\mathsf{SO}(3)$. Let us explain this. Given a real Lie algebra L of dimension n, we consider the *real* vector space $\mathbb{C}\otimes L$ of dimension $2n$. Here $\mathbb{C}$ is considered as two-dimensional real vector space and $\otimes$ is clearly also taken over $\mathbb{R}$. $\mathbb{C}\otimes L$ can be made into a real $2n$-dimensional Lie algebra by defining $[z_1\otimes X_1\,,\,z_2\otimes X_2] := z_1z_2\otimes[X_1,X_2]$ and $\mathbb{R}$-linear extension. This is easily checked to satisfy all axioms (180). The *complex double* of L is now defined to be the real Lie algebra $\mathbb{C}\otimes L$. For sure, $\mathbb{C}\otimes L$ has a natural complex structure, which allows to consider it as n-dimensional *complex* Lie algebra. In this case we[10] would call it $L^{\mathbb{C}}$, the *complexification* of L. However, we are interested in Lie algebras of Lie groups, which a priori are always considered as real (cf. Sect. A.9), regardless of the possible existence of a complex structure. Now let L be the Lie algebra of $\mathsf{SO}(3)$, i.e. $L=\mathrm{span}\{e_1,e_2,e_3\}$ where $[e_i,e_j]=\varepsilon_{ijk}e_k$. Consider $\mathbb{C}\otimes L$ and set $R_j := 1\otimes e_j$ and $cK_j := \mathrm{i}\otimes e_j$, so that $\mathbb{C}\otimes L=\mathrm{span}\{R_1,R_2,R_3,K_1,K_2,K_3\}$. In this basis the Lie brackets are just given by (51a-51c), showing that the homogeneous Lorentz Lie-algebra is indeed the complex double of the Lie algebra of $\mathsf{SO}(3)$.

The Lie algebra of the inhomogeneous Galilei group is formally obtained from (51) by taking the limit $1/c^2\to 0$, to that the right hand sides of (51c) and (51f) are now replaced with zero. This causes big structural changes. For example, the generators of boosts now generate an Abelian ideal in the homogeneous Galilei Lie-algebra (generated by R_i,K_i), implying that it is not even semisimple, whereas we just said that the homogeneous Lorentz Lie-algebra is simple. In the inhomogeneous Galilei Lie-algebra the K_i and P_i together generate an Abelian ideal. It is not perfect since the K_i and E do not occur on the right hand sides anymore.

One might argue that it is physically incorrect to take E/c^2 to zero in the limit $c\to\infty$. Rather, $E\to\infty$ as $c\to\infty$ since E contains a contribution mc^2 from the rest-energy of the system (m denotes the rest mass, which we wish to keep at a finite value). Hence, for an isolated system, one should rather set $E=mc^2+E_0$ and therefore have $E/c^2\to m$ in the limit as $c\to\infty$. Then the right hand side of (51c) is still zero in this limit but the right hand side of (51f) becomes proportional $m\,\delta_{ij}$, where m is now read as a new element of the Lie algebra that commutes with all other elements, i.e. lies in the center (in any irreducible representation it is therefore written as $m\mathbf{1}$ where $\mathbf{1}$ is the unit operator). Also, due to m being central, (51g) is maintained with E_0 replacing E.

The 11-dimensional Lie algebra so obtained is well known. It is a central extension of the inhomogeneous Galilei Lie-algebra, out of a unique 1-parameter family of inequivalent central extensions, labeled by the value of m. As is well known, it is this extension (and the corresponding 11-dimensional Lie group,

[10] The terminology used here is non-standard. Often the distinction between $\mathbb{C}\otimes L$ and $L^{\mathbb{C}}$ is not explicitly made, and even if it is, both are called "the complexification" of L.

sometimes called the Schrödinger group), which implement the Galilean symmetries in quantum mechanics by proper representations, whereas the inhomogeneous Galilei group only acts by ray-representations. Formally, the central element m then gives rise to superselection rules. There are certain analogs of this on the classical level; see [23].

The formal process by which the (inhomogeneous and homogeneous) Galilei Lie-algebra emerges from the Lorentz Lie-algebra is a special case of what is called a *contraction*, which was introduced in [31] just in order to understand precisely the way in which the Galilei Lie-algebra and group can be understood as limiting case of the Lorentz Lie-algebra and group respectively. The general idea can be briefly described as follows: Consider a Lie algebra L with decomposition into two linear subspaces $L = H \oplus H'$, none of which we a priori assume to be a Lie subalgebra. Choose an adapted basis $\{X_1, \cdots X_n, X'_1, \cdots X'_{n'}\}$ such that the unprimed elements span H and the primed elements H'. The Lie brackets have the general form

$$
\begin{aligned}
[X_a, X_b] &= C^c_{ab} X_c + C^{c'}_{ab} X'_{c'} , && \text{(52a)}\\
[X_a, X'_{b'}] &= C^c_{ab'} X_c + C^{c'}_{ab'} X'_{c'} , && \text{(52b)}\\
[X'_{a'}, X'_{b'}] &= C^c_{a'b'} X_c + C^{c'}_{a'b'} X'_{c'} . && \text{(52c)}
\end{aligned}
$$

We now rescale the primed generators, leaving the unprimed ones untouched,

$$
\begin{aligned}
X_a \mapsto Y_a &:= X_a , && \text{(53a)}\\
X'_{a'} \mapsto Y'_{a'} &:= \epsilon X'_{a'} && \text{(53b)}
\end{aligned}
$$

and write down (52) in terms of the new basis:

$$
\begin{aligned}
[Y_a, Y_b] &= C^c_{ab} Y_c + \epsilon^{-1} C^{c'}_{ab} Y'_{c'} , && \text{(54a)}\\
[Y_a, Y'_{b'}] &= \epsilon C^c_{ab'} Y_c + C^{c'}_{ab'} Y'_{c'} , && \text{(54b)}\\
[Y'_{a'}, Y'_{b'}] &= \epsilon^2 C^c_{a'b'} Y_c + \epsilon C^{c'}_{a'b'} Y'_{c'} . && \text{(54c)}
\end{aligned}
$$

We wish to formally take the limit $\epsilon \to 0$. Clearly this cannot be done unless the terms $\propto \epsilon^{-1}$ in (54a) all vanish, i.e. unless $C^{c'}_{ab} = 0$, which is equivalent to saying that $H := \text{span}\{X_1, \cdots X_n\}$ must be a Lie-subalgebra of L. Assuming that this is the case, the limit can be taken and the following Lie algebra emerges:

$$
\begin{aligned}
[Y_a, Y_b] &= C^c_{ab} Y_c , && \text{(55a)}\\
[Y_a, Y'_{b'}] &= C^{c'}_{ab'} Y'_{c'} , && \text{(55b)}\\
[Y'_{a'}, Y'_{b'}] &= 0 . && \text{(55c)}
\end{aligned}
$$

Thus we see that in the limit the subalgebra H survives whereas the linear space H' turns into an Abelian ideal. Hence the limit Lie algebra is a semi-direct sum of the original Lie subalgebra H with the Abelian ideal H'. It is called the *contraction of L over H*, since H stays intact and the rest is contracted. On the level of Lie groups one might think of the contracted group (the group

generated by H') as an infinitesimal neighborhood of the group one contracts over (the group generated by H) within the full Lie group (the group generated by L).

This applies to the transition Lorentz $\to$ Galilei as follows: In the homogeneous case, we decompose the Lorentz Lie-algebra into the Lie subalgebra $H = \mathrm{span}\{J_1, J_2, J_3\}$ and the linear subspace $H' = \mathrm{span}\{K_1, K_2, K_3\}$, and then contract it over H to obtain the homogeneous Galilei Lie-algebra. In the inhomogeneous case we set $H = \mathrm{span}\{J_1, J_2, J_3, E\}$, which is indeed a Lie subalgebra as seen from (51), and $H' = \mathrm{span}\{K_1, K_2, K_3, P_1, P_2, P_3\}$. Contracting over H then just results in making H' Abelian, i.e. annihilating the right hand sides of (51c) and (51f), which just results in the inhomogeneous Galilei Lie-algebra. Its structure as semi-direct sum with H' as Abelian ideal is just the Lie-algebra analog of the semi-direct product structure (36).

4.4 Composing Boosts

After this digression into Lie algebras we return to the level of groups. More specifically, we are now interested in the composition of two boosts, $B(\boldsymbol{\beta}_1)$ and $B(\boldsymbol{\beta}_2)$. The matrix product can be easily computed using the explicit form of $B(\boldsymbol{\beta})$ as given in (48). We set $\gamma_i := \gamma(\boldsymbol{\beta}_i)$ for $i = 1, 2$ and denote by $\boldsymbol{\beta}_{i\parallel}$ and $\boldsymbol{\beta}_{i\perp}$ the components of $\boldsymbol{\beta}_i$ parallel and perpendicular to the other velocity respectively. The angle between $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ is denote by φ, i.e. $\hat{\boldsymbol{\beta}}_1 \cdot \hat{\boldsymbol{\beta}}_2 = \cos\varphi$. Then the matrix product has the general form (39), where

$$\gamma = \gamma_1\gamma_1(1 + \boldsymbol{\beta}_1 \cdot \boldsymbol{\beta}_2) = \gamma_1\gamma_1(1 + \beta_1\beta_2\cos\varphi)\,, \tag{56a}$$

$$\mathbf{a} = \gamma_1\gamma_2\big(\boldsymbol{\beta}_2 + \boldsymbol{\beta}_{1\parallel} + \gamma_2^{-1}\boldsymbol{\beta}_{1\perp}\big)\,, \tag{56b}$$

$$\mathbf{b} = \gamma_1\gamma_2\big(\boldsymbol{\beta}_1 + \boldsymbol{\beta}_{2\parallel} + \gamma_1^{-1}\boldsymbol{\beta}_{2\perp}\big)\,, \tag{56c}$$

$$\begin{aligned}\mathbf{M} = {} & \mathbf{1}_3 + (\gamma_1 - 1)\,\hat{\boldsymbol{\beta}}_1 \otimes \hat{\boldsymbol{\beta}}_1^{\top} + (\gamma_2 - 1)\,\hat{\boldsymbol{\beta}}_2 \otimes \hat{\boldsymbol{\beta}}_2^{\top} \\ & + \big(\beta_1\gamma_1\beta_2\gamma_2 + (\gamma_1 - 1)(\gamma_2 - 1)\hat{\boldsymbol{\beta}}_1 \cdot \hat{\boldsymbol{\beta}}_2\big)\,\hat{\boldsymbol{\beta}}_1 \otimes \hat{\boldsymbol{\beta}}_2^{\top}\,.\end{aligned} \tag{56d}$$

The resulting boost and rotation parameters will be called $\boldsymbol{\beta} = \boldsymbol{\beta}_1 \star \boldsymbol{\beta}_2$ and $\mathbf{D} = \mathbf{T}[\boldsymbol{\beta}_1, \boldsymbol{\beta}_2]$ respectively. Hence we have:

$$B(\boldsymbol{\beta}_1) \cdot B(\boldsymbol{\beta}_2) = B(\boldsymbol{\beta}_1 \star \boldsymbol{\beta}_2) \cdot R(\mathbf{T}[\boldsymbol{\beta}_1, \boldsymbol{\beta}_2])\,. \tag{57}$$

The operation $\star$ entails the law of how to compose velocities in SR. $T[\boldsymbol{\beta}_1, \boldsymbol{\beta}_2]$ is called the "Thomas rotation". Its existence (i.e. it being non trivial) means that pure boosts do not form a subgroup in the Lorentz group, in contrast to the Galilei group.

The functional form of the $\star$ operation follows from (56), since $\boldsymbol{\beta}_1 \star \boldsymbol{\beta}_2 = \mathbf{b}/\gamma$:

$$\boldsymbol{\beta}_1 \star \boldsymbol{\beta}_2 = \frac{\boldsymbol{\beta}_1 + \boldsymbol{\beta}_{2\parallel} + \gamma_1^{-1}\boldsymbol{\beta}_{2\perp}}{1 + \boldsymbol{\beta}_1 \cdot \boldsymbol{\beta}_2}\,. \tag{58}$$

Comparing (56b) with (56c) shows $\mathbf{a}/\gamma = \boldsymbol{\beta}_2 \star \boldsymbol{\beta}_1$. Equation (45) then shows

$$\boldsymbol{\beta}_1 \star \boldsymbol{\beta}_2 = \mathbf{T}[\boldsymbol{\beta}_1, \boldsymbol{\beta}_2] \cdot (\boldsymbol{\beta}_2 \star \boldsymbol{\beta}_1)\,, \tag{59}$$

which in turn implies (we write $\mathbf{T}^{-1}[-,-]$ for the inverse matrix $(\mathbf{T}[-,-])^{-1}$)

$$\mathbf{T}[\boldsymbol{\beta}_1, \boldsymbol{\beta}_2] = \mathbf{T}^{-1}[\boldsymbol{\beta}_2, \boldsymbol{\beta}_1]\,. \tag{60}$$

Let now $\mathbf{D} \in \mathsf{SO}(3)$ be any rotation; then (58) shows that $\star$ obeys

$$(\mathbf{D} \cdot \boldsymbol{\beta}_1) \star (\mathbf{D} \cdot \boldsymbol{\beta}_2) = \mathbf{D} \cdot (\boldsymbol{\beta}_1 \star \boldsymbol{\beta}_2)\,, \tag{61}$$

which combined with (59) also shows that

$$\mathbf{T}[\mathbf{D} \cdot \boldsymbol{\beta}_1, \mathbf{D} \cdot \boldsymbol{\beta}_2] = \mathbf{D} \cdot \mathbf{T}[\boldsymbol{\beta}_1, \boldsymbol{\beta}_2] \cdot \mathbf{D}^{-1}\,. \tag{62}$$

The Thomas rotation takes place in the plane $\mathrm{span}\{\mathbf{a}, \mathbf{b}\} = \mathrm{span}\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2\}$. The cosine of the angle of rotation, θ, follows from (46c) and (56d):

$$\cos\theta = 1 - \frac{(\gamma_1 - 1)(\gamma_2 - 1)}{\gamma + 1} \sin^2\varphi\,, \tag{63a}$$

where we used (56a) to eliminate a term $\propto \cos\varphi$. It shows that $\mathbf{T}[\boldsymbol{\beta}_1, \boldsymbol{\beta}_2] = \mathbf{1}_3$ iff $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are either parallel ($\varphi = 0$) or anti-parallel ($\varphi = \pi$). We can now again make use of (56a) to eliminate γ in favor of γ_1, γ_2, and $\cos\varphi$, so as to make $\cos\theta$ a function of the moduli β_1, β_2 of the velocities and the angle φ between them:

$$\cos\theta = 1 - \frac{(\gamma_1 - 1)(\gamma_2 - 1)\sin^2\varphi}{1 + \gamma_1\gamma_2 + \sqrt{(\gamma_1^2 - 1)(\gamma_2^2 - 1)}\,\cos\varphi}\,. \tag{63b}$$

Alternatively we can use (56a) to express $\cos\theta$ as function of the tree moduli β_1, β_2, and $\beta = \|\boldsymbol{\beta}_1 \star \boldsymbol{\beta}_2\|$, which assumes a nice symmetric form:[11]

$$\cos\theta = \frac{(1 + \gamma + \gamma_1 + \gamma_2)^2}{(1 + \gamma)(1 + \gamma_1)(1 + \gamma_2)} - 1\,. \tag{63c}$$

Figure 1 illustrates the laws (58) and (59) of the Thomas rotation for a special case in which the two velocities are perpendicular. In such cases θ ranges between 0 and $\pi/2$, as can be immediately deduced from (63b). The sense of the Thomas rotation in the $\boldsymbol{\beta}_1\boldsymbol{\beta}_2$-plane is negative (we orient this plane in the usual way, such that $\boldsymbol{\beta}_1 \times \boldsymbol{\beta}_2$ defines the direction of the normal).

Generally θ ranges between 0 and π. More precisely, take fixed moduli β_1 and β_2 and consider $\cos\theta$ as function of φ as given by (63b). For $\varphi = 0$ and $\varphi = \pi$ this function has obvious maxima (where $\cos\theta = 1$) and hence must have a minimum inbetween, which corresponds to a *maximal* value of θ. Using (63b) we compute that this maximum of θ occurs at a value φ_m which obeys

$$\cos\varphi_m = -\sqrt{\frac{(\gamma_1 - 1)(\gamma_2 - 1)}{(\gamma_1 + 1)(\gamma_2 + 1)}}\,, \tag{64}$$

[11] This derivation, albeit straightforward, is a little tedious. A more elegant derivation, using Clifford algebra, is given in [47].

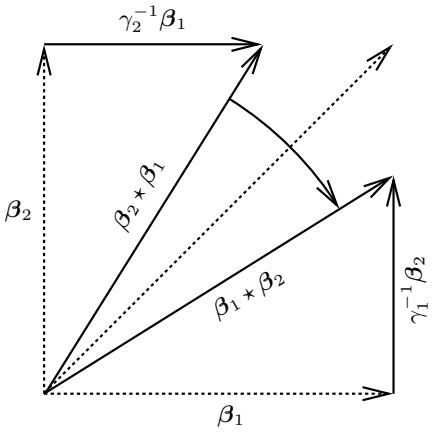


Fig. 1. Addition of perpendicular velocities $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ of modulus $\beta_1 = \beta_2 = 0.78$ so that $\gamma_1^{-1} = \gamma_2^{-1} = 5/8$. In this case (58) gives $\boldsymbol{\beta}_1 \star \boldsymbol{\beta}_2 = \boldsymbol{\beta}_1 + \frac{5}{8}\boldsymbol{\beta}_2$ and likewise $\boldsymbol{\beta}_2 \star \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 + \frac{5}{8}\boldsymbol{\beta}_1$. For comparison, the dashed arrow corresponds to the "classically" composed velocities (vector addition). According to (59), the rotation $\mathbf{T}[\boldsymbol{\beta}_1, \boldsymbol{\beta}_2]$ turns $\boldsymbol{\beta}_2 \star \boldsymbol{\beta}_1$ into $\boldsymbol{\beta}_1 \star \boldsymbol{\beta}_2$, as indicated by the *curved arrow*

(the negative sign shows that $\varphi_m > \pi/2$) and that the maximal value θ_m obeys

$$\cos\theta_m = 1 - 2\,\frac{(\gamma_1 - 1)(\gamma_2 - 1)}{(\gamma_1 + 1)(\gamma_2 + 1)} = -\cos(2\varphi_m)\,. \tag{65}$$

Hence we see that θ becomes larger than $\pi/2$ for sufficiently large values of γ_1 and γ_2. For example, if $\beta_1 = \beta_2 = \beta$, i.e. $\gamma_1 = \gamma_2 = \gamma$, the value of β above which θ_m exceeds $\pi/2$ is given by $2^{5/4}/(2^{1/2} + 2) \approx 0.985$. Equation (65) also shows that θ_m approaches its maximal value, π, only if γ_1 *and* γ_2 tend to infinity. In general, (63b) shows that in that limit $\cos\theta$ approaches $\cos\varphi$, which means that θ approaches $2\pi - \varphi$, since the Thomas rotation is in the negative sense relative to the orientation of the $\boldsymbol{\beta}_1$-$\boldsymbol{\beta}_2$ plane.

Finally, using (49) and (57), we can now write down the general composition law for Lorentz transformations:

$$L(\boldsymbol{\beta}_1, \mathbf{D}_1) \cdot L(\boldsymbol{\beta}_2, \mathbf{D}_2) = L\big(\boldsymbol{\beta}_1 \star \mathbf{D}_1 \cdot \boldsymbol{\beta}_2\,,\, \mathbf{T}[\boldsymbol{\beta}_1, \mathbf{D}_1 \cdot \boldsymbol{\beta}_2] \cdot \mathbf{D}_1 \cdot \mathbf{D}_2\big)\,. \tag{66}$$

Moreover, noting that $(B(\boldsymbol{\beta}))^{-1} = B(-\boldsymbol{\beta})$, equations (41,49) also show that

$$\big(L(\boldsymbol{\beta}, \mathbf{D})\big)^{-1} = L(-\,\mathbf{D}^{-1} \cdot \boldsymbol{\beta}\,,\, \mathbf{D}^{-1})\,. \tag{67}$$

Note that (67) and (34) are just the same analytic operations on the parameter spaces. The multiplication law (66) now replaces the semi-direct product structure (33) of the Galilei group, into which it turns in the limit $c \to \infty$. Indeed, writing $\boldsymbol{\beta} = \mathbf{v}/c$, the operation $\star$ between the $\mathbf{v}$'s approaches $+$ and the Thomas rotation $\mathbf{T}[-,-]$ becomes the identity, as one e.g. sees from (63b) for $\gamma_1, \gamma_2 \to 1$.

4.5 The Algebraic Structure of Velocity Composition

Let us say a little more about the algebraic structure behind (58). First of all, $\star$ defines a map

$$\star : \mathcal{B}_1 \times \mathcal{B}_1 \to \mathcal{B}_1\,, \quad (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \mapsto \boldsymbol{\beta}_1 \star \boldsymbol{\beta}_2\,, \tag{68}$$

where

$$\mathcal{B}_1 := \{\boldsymbol{\beta} \in \mathbb{R}^3 \;\mid\; \|\boldsymbol{\beta}\| < 1\} \tag{69}$$

is the open ball in 3-dimensional Euclidean space (here space of velocities/c). That its image lies indeed in $\mathcal{B}_1 \subset \mathbb{R}^3$ follows from (56a), which e.g. implies $\gamma < 2\gamma_1\gamma_2$. Hence $\star$ makes $\mathcal{B}_\infty$ into a *groupoid* (see below). Moreover, for each $\boldsymbol{\beta} \in \mathcal{B}_1$, we have

$$\mathbf{0} \star \boldsymbol{\beta} = \boldsymbol{\beta} \star \mathbf{0} = \boldsymbol{\beta}\,, \tag{70}$$

so that $\mathbf{0}$ is a unit with respect to $\star$. Each element also has an inverse (left and right):

$$\boldsymbol{\beta} \star (-\boldsymbol{\beta}) = (-\boldsymbol{\beta}) \star \boldsymbol{\beta} = \mathbf{0}\,. \tag{71}$$

We already saw in (59) that the Thomas rotation obstructs commutativity of $\star$. We now show that it also obstructs associativity. Consider the composition of three boosts $B(\boldsymbol{\beta}_1)\cdot B(\boldsymbol{\beta}_2)\cdot B(\boldsymbol{\beta}_3)$ and use associativity of matrix multiplication:

$$B(\boldsymbol{\beta}_1)\cdot\big(B(\boldsymbol{\beta}_2)\cdot B(\boldsymbol{\beta}_3)\big) = \big(B(\boldsymbol{\beta}_1)\cdot B(\boldsymbol{\beta}_2)\big)\cdot B(\boldsymbol{\beta}_3)\,. \tag{72a}$$

Iterated application of (57) shows that the left hand side is equal to

$$B\big(\boldsymbol{\beta}_1 \star (\boldsymbol{\beta}_2 \star \boldsymbol{\beta}_3)\big)\cdot R(\mathbf{T}[\boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \star \boldsymbol{\beta}_3]\cdot \mathbf{T}[\boldsymbol{\beta}_2, \boldsymbol{\beta}_3])\,, \tag{72b}$$

whereas the right hand side equals (also making use of (49)),

$$B\big((\boldsymbol{\beta}_1 \star \boldsymbol{\beta}_2) \star (\mathbf{T}[\boldsymbol{\beta}_1, \boldsymbol{\beta}_2]\cdot\boldsymbol{\beta}_3)\big)\cdot R\big(\mathbf{T}\big[\boldsymbol{\beta}_1 \star \boldsymbol{\beta}_2\,,\, \mathbf{T}[\boldsymbol{\beta}_1, \boldsymbol{\beta}_2]\cdot\boldsymbol{\beta}_3\big]\cdot \mathbf{T}[\boldsymbol{\beta}_1, \boldsymbol{\beta}_2]\big)\,. \tag{72c}$$

Expressions (72b) and (72c) are in polar decomposed form. Uniqueness then implies equality of the boost and rotation factors separately. For the boosts this implies

$$\boldsymbol{\beta}_1 \star (\boldsymbol{\beta}_2 \star \boldsymbol{\beta}_3) = (\boldsymbol{\beta}_1 \star \boldsymbol{\beta}_2) \star \big(\mathbf{T}[\boldsymbol{\beta}_1, \boldsymbol{\beta}_2]\cdot\boldsymbol{\beta}_3\big)\,, \tag{73}$$

which shows how the Thomas rotation obstructs associativity. The general identity obtained from equating the rotational parts of (72b) and (72c) does not interest us here. Rather, we wish to consider the special case where $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_3$. Then the product (72) is a symmetric and positive definite matrix[12], that is, it is a pure boost and therefore (trivially) polar decomposed. Hence the rotational part in (72b) must be the identity. This gives

$$\mathbf{T}[\boldsymbol{\beta}_1, \boldsymbol{\beta}_2] = \mathbf{T}[\boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \star \boldsymbol{\beta}_1] = \mathbf{T}[\boldsymbol{\beta}_1 \star \boldsymbol{\beta}_2, \boldsymbol{\beta}_2]\,, \tag{74}$$

[12] For matrices it is generally true that if B is positive definite and A invertible, then $A\cdot B\cdot A^t$ is again positive definite. Note that here A^t is the adjoint of A with respect to the Euclidean inner product.

where the second equality follows from the first by simultaneously taking the inverse and exchanging $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ (which leaves $\mathbf{T}[\boldsymbol{\beta}_1, \boldsymbol{\beta}_2]$ invariant according to (60)).

Now consider the following equation in $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$, and $\boldsymbol{\beta}_3$:

$$\boldsymbol{\beta}_1 \star \boldsymbol{\beta}_2 = \boldsymbol{\beta}_3 \,. \tag{75a}$$

Can we (uniquely) solve it for $\boldsymbol{\beta}_1$ given $\boldsymbol{\beta}_2$ and $\boldsymbol{\beta}_3$, or for $\boldsymbol{\beta}_2$ given $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_3$? Since each $\boldsymbol{\beta}$ has an inverse, associativity would immediately answer this in the affirmative.[13] But associativity fails to hold. However, the answer is still affirmative:

Proposition 3. *The unique solutions of (75a) for $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are given by*

$$\boldsymbol{\beta}_1 \;=\; \boldsymbol{\beta}_3 \star (-\mathbf{T}[\boldsymbol{\beta}_3, \boldsymbol{\beta}_2] \cdot \boldsymbol{\beta}_2) \,, \tag{75b}$$

$$\boldsymbol{\beta}_2 \;=\; (-\boldsymbol{\beta}_1) \star \boldsymbol{\beta}_3 \,. \tag{75c}$$

Proof. (75c) immediately follows from $\star$-multiplying (75a) with $-\boldsymbol{\beta}_1$ from the left and using (73), taking into account that $\mathbf{T}[-\boldsymbol{\beta}_1, \boldsymbol{\beta}_1] = \mathbf{1}_3$. The proof of (75b) is more difficult. One way that is not just "guessing and verifying", but rather arrives at the solution in a more systematic fashion, is to go back to the group level and consider the corresponding equation

$$L(\boldsymbol{\beta}_1, \mathbf{D_1}) \cdot L(\boldsymbol{\beta}_2, \mathbf{D}_2) = L(\boldsymbol{\beta}_3, \mathbf{D}_3) \,, \tag{76a}$$

whose parameter form is

$$\boldsymbol{\beta}_3 \;=\; \boldsymbol{\beta}_1 \star \mathbf{D_1} \cdot \boldsymbol{\beta}_2 \,, \tag{76b}$$

$$\mathbf{D_3} \;=\; \mathbf{T}[\boldsymbol{\beta}_1, \mathbf{D}_1 \cdot \boldsymbol{\beta}_2] \cdot \mathbf{D}_1 \cdot \mathbf{D}_2 \,. \tag{76c}$$

The group structure now tells us that the unique solution for $L(\boldsymbol{\beta}_1, \mathbf{D}_1)$ is, using (67),

$$L(\boldsymbol{\beta}_1, \mathbf{D}_1) = L(\boldsymbol{\beta}_3, \mathbf{D}_3) \cdot L(-\mathbf{D}_2^{-1} \cdot \boldsymbol{\beta}_2, \mathbf{D}_2^{-1}) \tag{77a}$$

whose parameter form is

$$\boldsymbol{\beta}_1 \;=\; \boldsymbol{\beta}_3 \star (-\mathbf{D}_3 \cdot \mathbf{D}_2^{-1} \cdot \boldsymbol{\beta}_2) \,, \tag{77b}$$

$$\mathbf{D}_1 \;=\; \mathbf{T}[\boldsymbol{\beta}_3 \,, -\mathbf{D}_3 \cdot \mathbf{D}_2^{-1} \cdot \boldsymbol{\beta}_2] \cdot \mathbf{D}_3 \cdot \mathbf{D}_2^{-1} \,. \tag{77c}$$

Due to the group structure (76) and (77) are equivalent. In particular, (77) is a consequence of (76). We now specialize to the case $\mathbf{D}_1 = \mathbf{1}_3$, in which (76b) just becomes (75a). Equation (76c) then becomes

$$\begin{aligned} \mathbf{D}_3 \cdot \mathbf{D}_2^{-1} \;&=\; \mathbf{T}[\boldsymbol{\beta}_1, \boldsymbol{\beta}_2] \\ &=\; \mathbf{T}[\boldsymbol{\beta}_1 \star \boldsymbol{\beta}_2 \,, \boldsymbol{\beta}_2] \qquad \text{using (74)} \\ &=\; \mathbf{T}[\boldsymbol{\beta}_3, \boldsymbol{\beta}_2] \qquad \text{using (76b)} \,. \end{aligned} \tag{78}$$

Inserting this into (77b) gives (75b). □

[13] For then we could e.g. $\star$-multiply (75a) with $-\boldsymbol{\beta}_2$ from the right and get on the left hand side $(\boldsymbol{\beta}_1 \star \boldsymbol{\beta}_2) \star (-\boldsymbol{\beta}_2) = \boldsymbol{\beta}_1 \star (\boldsymbol{\beta}_2 \star (-\boldsymbol{\beta}_2)) = \boldsymbol{\beta}_1$.

Let us relate these findings to some algebraic terminology. A *groupoid* is a set S with some map $\phi : S \times S \to S$. Hence $\star$ makes the open unit ball $\mathcal{B}_1 \subset \mathbb{R}^3$ into a groupoid. An associative groupoid is called a *semigroup* (so we don't have a semigroup) . A groupoid S is called a *quasigroup* if for any pair $(a, b) \in S \times S$ there is a unique pair $(x, y) \in S \times S$ such that $\phi(x, a) = b$ and $\phi(a, y) = b$. In our case we have just seen that the unique pair (x, y) associated to $(a, b) = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ is $x = \boldsymbol{\beta}_2 \star (-\mathbf{T}[\boldsymbol{\beta}_2, \boldsymbol{\beta}_1] \cdot \boldsymbol{\beta}_1)$ and $y = (-\boldsymbol{\beta}_1) \star \boldsymbol{\beta}_2$. If a common unit element exists, as in (70), one calls it a *quasigroup with unit* or simply a *loop*. Note that in this case the existence of a unique inverse for each element follows. In some sense a loop is as close as you can get to the structure of a group if you drop associativity. This is the algebraic structure of velocity space in SR. Much original work on this has been done by A. Ungar, starting with [44], where e.g. the precise way in which strict associativity fails (i.e. (73)) was first spelled out; see also his comprehensive treatise [45] and references therein. In a more recent book [46] the same author systematically develops the intimate relation to hyperbolic geometry. A brief history of the research on these generalized algebraic structures is given in [40].

Let us briefly come back to the composition formulae (75). We interpret $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, and $\boldsymbol{\beta}_3$ as velocities of frames: $\boldsymbol{\beta}_1$ is the velocity of frame 2 with respect to (i.e. measured in) frame 1. $\boldsymbol{\beta}_2$ is the velocity of frame 3 with respect to frame 2. Finally, $\boldsymbol{\beta}_3$ is the velocity of frame 3 with respect to frame 1. Then, using (56a), it is easy to derive the following expressions for the moduli of $\boldsymbol{\beta}_3$ and $\boldsymbol{\beta}_2$:

$$\beta_3^2 = \frac{(\boldsymbol{\beta}_1 + \boldsymbol{\beta}_2)^2 - (\boldsymbol{\beta}_1 \times \boldsymbol{\beta}_2)^2}{(1 + \boldsymbol{\beta}_1 \cdot \boldsymbol{\beta}_2)^2}\,, \tag{79a}$$

$$\beta_2^2 = \frac{(\boldsymbol{\beta}_3 - \boldsymbol{\beta}_1)^2 - (\boldsymbol{\beta}_3 \times \boldsymbol{\beta}_1)^2}{(1 - \boldsymbol{\beta}_3 \cdot \boldsymbol{\beta}_1)^2}\,. \tag{79b}$$

β_2 is the modulus of the relative velocity between frames 2 and 3 as function of the velocities of these frames with respect to a third one (here frame 1). It may either be interpreted as velocity of frame 3 with respect to frame 2, (as above) or as velocity of frame 2 with respect to frame 3 (reciprocity of frame velocities, see Sect. 3). Accordingly, the right hand side of (79b) is symmetric under the exchange $\boldsymbol{\beta}_1 \leftrightarrow \boldsymbol{\beta}_3$.

4.6 The Geometric Structure of Velocity Composition

Even though the discussion of the geometry behind velocity composition belongs, strictly speaking, to the next, the geometry section, it is so intimately related to the discussion just given that it seems more appropriate to place the two right next to each other.

More precisely, the composition law for velocities is intimately related with hyperbolic geometry (i.e. geometry on spaces with constant negative curvature), as was first pointed out by Sommerfeld [41], Varičak [49, 50], Robb [39], and Borel [11]. More recently the subject was elaborated on by Ungar [46]. The general reason is that the space of four-velocities

$$\mathcal{H}_c := \{u \in \mathbb{R}^4 \mid g(u,u) = c^2\} \subset \mathbb{R}^4\,, \tag{80}$$

is a 3-dimensional hyperbola in $(\mathbb{R}^4, g)$, whose induced metric (g restricted to the tangent bundle of $\mathcal{H}_c$) is of constant negative curvature.

The space of velocities is parameterized by $\boldsymbol{\beta} \in \mathcal{B}_1$ (cf. (69)). Formula (79b) may now be read as endowing $\mathcal{B}_1$ with a distance function. From it we can read off the Riemannian metric by just applying it to two infinitesimally nearby velocities $\boldsymbol{\beta} = \boldsymbol{\beta}_1$ and $\boldsymbol{\beta} + d\boldsymbol{\beta} = \boldsymbol{\beta}_3$. Then β_2^2 gives us the square of their distance, ds^2, and we obtain

$$ds^2 = \frac{d\boldsymbol{\beta}^2 - (\boldsymbol{\beta} \times d\boldsymbol{\beta})^2}{\big(1-\beta^2\big)^2} \tag{81a}$$

$$= \frac{d\beta^2}{\big(1-\beta^2\big)^2} + \frac{\beta^2}{1-\beta^2}\left(d\theta^2 + \sin^2\theta\, d\varphi^2\right) \tag{81b}$$

$$= \frac{dr^2}{1+r^2} + r^2\left(d\theta^2 + \sin^2\theta\, d\varphi^2\right). \tag{81c}$$

Here (β, θ, φ) are just the ordinary spherical polar coordinates in $\boldsymbol{\beta}$-space and $r := \beta/\sqrt{1-\beta^2}$. This is easily recognized as Riemannian metric of constant negative curvature, e.g. by comparing (81c) with the spatial part of the $k = -1$ standard FRW-metric in cosmology. The geodesic between the origin and a point (β, θ, φ) is just the radial segment, whose length is

$$s = \int_0^\beta \frac{d\beta'}{\sqrt{1-\beta'^2}} = \tanh^{-1}\beta\,. \tag{82}$$

Hence the rapidity (26) turns out to be just the geodesic distance in velocity space. This explains why in terms of it the composition of velocities in the same direction is just ordinary addition; compare the remark following equation (26).

In this geometric setting the law (79a) for the modulus of the composed velocities just turns into the law for the length of the third side of a geodesic triangle as function of the length of the two other sides and the angle between them. This is most easily read off from (56a) if rewritten in terms of rapidities, i.e. $\gamma_i = \cosh\rho_i$ and $\beta_i\gamma_i = \sinh\rho_i$:

$$\cosh\rho_3 = \cosh\rho_1\,\cosh\rho_1 + \sinh\rho_1\,\sinh\rho_2\,\cos\varphi\,. \tag{83}$$

This is just the well known "cosine-law" for hyperbolic triangles, the connection of which with the law of composing velocities in SR was first pointed out by Sommerfeld [41] and later, independently, by Borel [11].

A beautiful application of the hyperbolic geometry of velocity space (80) concerns Thomas rotation [47]. Suppose a torque-free gyro is carried along the worldline $z(\tau)$ of an observer. The *hodograph* is the curve $\dot{z}(\tau)$ on $\mathcal{H}_c$ and $\dot{z}(\tau)^\perp$ can be identified with the tangent plane to $\mathcal{H}_c$ at $\dot{z}(\tau)$. At each instant the gyro's angular-momentum vector lies in this tangent plane and along the worldline it

is Fermi-Walker transported. We recall that given a vector field X along the worldline z, the Fermi-Walker derivative of X along z is defined by

$$F_{\dot{z}}X := (\nabla_{\dot{z}}X_{\parallel})_{\parallel} + (\nabla_{\dot{z}}X_{\perp})_{\perp}\,, \tag{84}$$

where $\parallel$ and $\perp$ denote the g-orthogonal projections parallel and perpendicular to the worldline's tangent direction $\dot{z}$. Applied to the gyro's angular momentum vector one sees that the law of Fermi-Walker transportation along z turns into the law of parallel propagation along the hodograph on $\mathcal{H}_c$ with respect to the Levi-Civita connection for the hyperbolic metric that $\mathcal{H}_c$ inherits from its embedding into Minkowski space.[14] Applied to spatially periodic orbits the holonomy of their closed hodographs in the tangent bundle of $\mathcal{H}_c$ is then just Thomas' rotation. This neat geometric idea goes back to Borel [11], who sketched it almost 15 years before Thomas' paper [42] appeared.

5 Geometric Structures in Minkowski Space

5.1 Preliminaries

Let us generally consider n dimensional Minkowski space $\mathbb{M}^n$, that is, the affine space over an n-dimensional, real vector space V with a non-degenerate bilinear form g of signature $(1, n-1)$ (compare Sect. A.7 and Sect. A.2 respectively). We introduce the following notations:

$$v \cdot w \;:=\; g(v,w) \qquad \text{and} \qquad \|v\|_g := \sqrt{|g(v,v)|}\,. \tag{85}$$

We shall also simply write v^2 for $v \cdot v$. A vector $v \in V$ is called *timelike, lightlike*, or *spacelike* according to v^2 being > 0, $= 0$, or < 0 respectively. Non-spacelike vectors are also called *causal* and their set, $\bar{\mathcal{C}} \subset V$, is called the *causal-doublecone*. Its interior, $\mathcal{C}$, is called the *chronological-doublecone* and its boundary, $\mathcal{L}$, the *light-doublecone*:

$$\bar{\mathcal{C}} := \{v \in V \mid v^2 \geq 0\}\,, \tag{86a}$$

$$\mathcal{C} := \{v \in V \mid v^2 > 0\}\,, \tag{86b}$$

$$\mathcal{L} := \{v \in V \mid v^2 = 0\}\,. \tag{86c}$$

A linear subspace $V' \subset V$ is called timelike, lightlike, or spacelike according to $g\big|_{V'}$ being indefinite, negative semi-definite but not negative definite, or negative definite respectively. Instead of the usual Cauchy-Schwarz-inequality we have

$$v^2 w^2 \;\leq\; (v \cdot w)^2 \quad \text{for span}\{v,w\} \text{ timelike}\,, \tag{87a}$$

$$v^2 w^2 \;=\; (v \cdot w)^2 \quad \text{for span}\{v,w\} \text{ lightlike}\,, \tag{87b}$$

$$v^2 w^2 \;\geq\; (v \cdot w)^2 \quad \text{for span}\{v,w\} \text{ spacelike}\,. \tag{87c}$$

[14] Generally, the Levi-Civita covariant derivative of a submanifold is obtained from the (covariant) derivative of the ambient manifold by restricting it to tangent vectors and subsequently projecting the result tangentially to the submanifold.

Given a set $W \subset V$ (not necessarily a subspace[15]), its g-orthogonal complement is the subspace

$$W^\perp := \{v \in V \mid v \cdot w = 0, \forall w \in W\}\,. \tag{88}$$

If $v \in V$ is lightlike then $v \in v^\perp$. In fact, $v^\perp$ is the unique lightlike hyperplane containing v. On the other hand, if v is timelike/spacelike $v^\perp$ is spacelike/timelike and $v \notin v^\perp$.

Given any subset $W \subset V$, we can attach it to a point p in $\mathbb{M}^n$:

$$W_p := p + W := \{p + w \mid w \in W\}\,. \tag{89}$$

In particular, the causal-, chronological-, and light-doublecones at $p \in \mathbb{M}^n$ are given by:

$$\bar{\mathcal{C}}_p := p + \bar{\mathcal{C}}\,, \tag{90a}$$
$$\mathcal{C}_p := p + \mathcal{C}\,, \tag{90b}$$
$$\mathcal{L}_p := p + \mathcal{L}\,. \tag{90c}$$

If W is a subspace of V then W_p is an affine subspace of $\mathbb{M}^n$ over W. If W is time-, light-, or spacelike then W_p is also called time-, light-, or spacelike. Of particular interest are the hyperplanes $v_p^\perp$ which are timelike, lightlike, or spacelike according to v being spacelike, lightlike, or timelike respectively.

Two points $p, q \in \mathbb{M}^n$ are said to be timelike-, lightlike-, or spacelike separated if the line joining them (equivalently: the vector $p - q$) is timelike, lightlike, or spacelike respectively. Non-spacelike separated points are also called causally separated and the line though them is called a causal line.

It is easy to show that the relation $v \sim w \Leftrightarrow v \cdot w > 0$ defines an equivalence relation on the set of timelike vectors. (Only transitivity is non-trivial, i.e. if $u \cdot v > 0$ and $v \cdot w > 0$ then $u \cdot w > 0$. To show this, decompose u and w into their components parallel and perpendicular to v.). Each of the two equivalence classes is a *cone* in V, that is, closed under addition and multiplication with positive numbers. Vectors in the same class are said to have the same time orientation. In the same fashion the relation $v \sim w \Leftrightarrow v \cdot w \geq 0$ defines an equivalence relation on the set of causal vectors, with both equivalence classes being again cones. The existence of these equivalence relations is expressed by saying that $\mathbb{M}^n$ is *time orientable*. Picking one of the two possible time orientations is then equivalent to specifying a single timelike reference vector, v_*, whose equivalence class of directions may be called the *future*. This being done we can speak of the future (or forward) $(+)$ and past (or backward) $(-)$ cones:

$$\bar{\mathcal{C}}^\pm := \{v \in \bar{\mathcal{C}} \mid v \cdot v_* \gtrless 0\}\,, \tag{91a}$$
$$\mathcal{C}^\pm := \{v \in \bar{\mathcal{C}} \mid v \cdot v_* \gtrless 0\}\,, \tag{91b}$$
$$\mathcal{L}^\pm := \{v \in \bar{\mathcal{L}} \mid v \cdot v_* \gtrless 0\}\,. \tag{91c}$$

[15] By a 'subspace' of a vector space we always understand a sub vector-space.

Note that $\bar{\mathcal{C}}^{\pm} = \mathcal{C}^{\pm} \cup \mathcal{L}^{\pm}$ and $\mathcal{C}^{\pm} \cap \mathcal{L}^{\pm} = \emptyset$. Usually $\mathcal{L}^{+}$ is called the future and $\mathcal{L}^{-}$ the past lightcone. Mathematically speaking this is an abuse of language since, in contrast to $\bar{\mathcal{C}}^{\pm}$ and $\mathcal{C}^{\pm}$, they are not cones: They are each invariant (as sets) under multiplication with positive real numbers, but adding to vectors in $\mathcal{L}^{\pm}$ will result in a vector in $\mathcal{C}^{\pm}$ unless the vectors were parallel.

As before, these cones can be attached to the points in $\mathbb{M}^n$. We write in a straightforward manner:

$$\bar{\mathcal{C}}_p^{\pm} := p + \bar{\mathcal{C}}^{\pm}\,, \tag{92a}$$

$$\mathcal{C}_p^{\pm} := p + \mathcal{C}^{\pm}\,, \tag{92b}$$

$$\mathcal{L}_p^{\pm} := p + \mathcal{L}^{\pm}\,. \tag{92c}$$

The Cauchy-Schwarz inequalities (87) result in various cases for generalized triangle inequalities. Clearly, for spacelike vectors, one just has the ordinary triangle inequality. But for causal or timelike vectors one has to distinguish the cases according to the relative time orientations. For example, for timelike vectors of equal time orientation, one obtains the reversed triangle inequality:

$$\|v + w\|_g \geq \|v\|_g + \|w\|_g\,, \tag{93}$$

with equality iff v and w are parallel. It expresses the geometry behind the "twin paradox".

Before we turn to the next section, we remark that any bijective map $\phi : \mathbb{M}^n \to \mathbb{M}^n$ that satisfies $d(p,q) = d(\phi(p), \phi(q))$, where $d(p,q) := \|p - q\|_g$, is necessarily affine linear. This follows immediately from the corresponding statement for vector spaces, as given in Proposition 9. The results in the following section should be considered as strengthenings of this statement.

5.2 Causality Relations and the Lorentz Group

The family of cones $\{\bar{\mathcal{C}}_q^{+} \mid q \in \mathbb{M}^n\}$ defines a partial order relation, denoted by $\geq$ (cf. Sect. A.1), on spacetime as follows: $p \geq q$ iff $p \in \bar{\mathcal{C}}_q^{+}$, i.e. iff $p - q$ is causal and future pointing. Similarly, the family $\{\mathcal{C}_q^{+} \mid q \in \mathbb{M}^n\}$ defines a strict partial order, denoted by $>$ (cf. Sect. A.1): $p > q$ iff $p \in \mathcal{C}_q^{+}$, i.e. if $p - q$ is timelike and future pointing. There is a third relation, called $\succcurlyeq$, defined as follows: $p \succcurlyeq q$ iff $p \in \mathcal{L}_q^{+}$, i.e. p is on the future lightcone at q. It is not a partial order due to the lack of transitivity, which, in turn, is due to the lack of the lightcone being a cone (in the proper mathematical sense explained above). Replacing the future $(+)$ with the past $(-)$ cones gives the relations $\leq$, $<$, and $\preccurlyeq$.

It is obvious that the action of $\mathsf{ILor}^{\uparrow}$ (spatial reflections are permitted) on $\mathbb{M}^n$ maps each of the six families of cones (92) into itself and therefore leave each of the six relations invariant. For example: Let $p > q$ and $f \in \mathsf{ILor}^{\uparrow}$, then $(p-q)^2 > 0$ and $p-q$ future pointing, but also $(f(p)-f(q))^2 > 0$ and $f(p)-f(q)$ future pointing, hence $f(p) > f(q)$. Another set of "obvious" transformations of $\mathbb{M}^n$ leaving these relations invariant is given by all dilations:

$$d_{(\lambda,m)} : \mathbb{M}^n \to \mathbb{M}^n\,, \quad p \mapsto d_{(\lambda,m)}(p) := \lambda(p-m)+m\,, \tag{94}$$

where $\lambda \in \mathbb{R}_+$ is the constant dilation-factor and $m \in \mathbb{M}^n$ the center. This follows from $\big(d_{\lambda,m}(p) - d_{\lambda,m}(q)\big)^2 = \lambda^2(p-q)^2$, $\big(d_{\lambda,m}(p) - d_{\lambda,m}(q)\big)\cdot v_* = \lambda(p-q)\cdot v_*$, and the positivity of λ. Since translations are already contained in $\mathsf{ILor}^\uparrow$, the group generated by $\mathsf{ILor}^\uparrow$ and all $d_{\lambda,m}$ is the same as the group generated by $\mathsf{ILor}^\uparrow$ and all $d_{\lambda,m}$ for fixed m.

A seemingly difficult question is this: What are the most general transformations of $\mathbb{M}^n$ that preserve those relations? Here we understand "transformation" synonymously with "bijective map", so that each transformation f has in inverse f^{-1}. "Preserving the relation" is taken to mean that f *and* f^{-1} preserve the relation. Then the somewhat surprising answer to the question just posed is that, in three or more spacetime dimensions, there are no other such transformations besides those already listed:

Theorem 1. *Let $\succ$ stand for any of the relations $\geq, >, \gg$ and let f be a bijection of $\mathbb{M}^n$ with $n \geq 3$, such that $p \succ q$ implies $f(p) \succ f(q)$ and $f^{-1}(p) \succ f^{-1}(q)$. Then f is the composition of an Lorentz transformation in $\mathsf{ILor}^\uparrow$ with a dilation.*

Proof. These results were proven by A.D. Alexandrov and independently by E.C. Zeeman. A good review of Alexandrov's results is [1]; Zeeman's paper is [51]. The restriction to $n \geq 3$ is indeed necessary, as for $n = 2$ the following possibility exists: Identify $\mathbb{M}^2$ with $\mathbb{R}^2$ and the bilinear form $g(z,z) = x^2 - y^2$, where $z = (x,y)$. Set $u := x - y$ and $v := x + y$ and define $f : \mathbb{R}^2 \to \mathbb{R}^2$ by $f(u,v) := (h(u), h(v))$, where $h : \mathbb{R} \to \mathbb{R}$ is any smooth function with $h' > 0$. This defines an orientation preserving diffeomorphism of $\mathbb{R}^2$ which transforms the set of lines $u =$ const. and $v =$ const. respectively into each other. Hence it preserves the families of cones (92a). Since these transformations need not be affine linear they are not generated by dilations and Lorentz transformations. □

These results may appear surprising since without a continuity requirement one might expect all sorts of wild behavior to allow for more possibilities. However, a little closer inspection reveals a fairly obvious reason for why continuity is implied here. Consider the case in which a transformation f preserves the families $\{\mathcal{C}^+_q \mid q \in \mathbb{M}^n\}$ and $\{\mathcal{C}^-_q \mid q \in \mathbb{M}^n\}$. The open diamond-shaped sets (usually just called "open diamonds"),

$$U(p,q) := (\mathcal{C}^+_p \cap \mathcal{C}^-_q) \cup (\mathcal{C}^+_q \cap \mathcal{C}^-_p)\,, \tag{95}$$

are obviously open in the standard topology of $\mathbb{M}^n$ (which is that of $\mathbb{R}^n$). Note that at least one of the intersections in (95) is always empty. Conversely, is is also easy to see that each open set of $\mathbb{M}^n$ contains an open diamond. Hence the topology that is defined by taking the $U(p,q)$ as subbase (the basis being given by their finite intersections) is equivalent to the standard topology of $\mathbb{M}^n$. But, by hypothesis, f and f^{-1} preserves the cones $\mathcal{C}^\pm_q$ and therefore open sets, so that f must, in fact, be a homeomorphism.

There is no such *obvious* continuity input if one makes the strictly weaker requirement that instead of the cones (92) one only preserves the doublecones (90). Does that allow for more transformations, except for the obvious time reflection? The answer is again in the negative. The following result was shown by Alexandrov (see his review [1]) and later, in a different fashion, by Borchers and Hegerfeld [10]:

Theorem 2. *Let $\sim$ denote any of the relations: $p \sim q$ iff $(p-q)^2 \geq 0$, $p \sim q$ iff $(p-q)^2 > 0$, or $p \sim q$ iff $(p-q)^2 = 0$. Let f be a bijection of $\mathbb{M}^n$ with $n \geq 3$, such that $p \sim q$ implies $f(p) \sim f(q)$ and $f^{-1}(p) \sim f^{-1}(q)$. Then f is the composition of an Lorentz transformation in ILor with a dilation.*

All this shows that, up to dilations, Lorentz transformations can be characterized by the causal structure of Minkowski space. Let us focus on a particular subcase of Theorem 2, which says that any bijection f of $\mathbb{M}^n$ with $n \geq 3$, which satisfies $\|p-q\|_g = 0 \Leftrightarrow \|f(p) - f(q)\|_g = 0$ must be the composition of a dilation and a transformation in ILor. This is sometimes referred to as *Alexandrov's theorem*. It is, to my knowledge, the closest analog in Minkowskian geometry to the famous theorem of Beckman and Quarles [4], which refers to Euclidean geometry and reads as follows:[16]

Theorem 3 (Beckman and Quarles 1953). *Let $\mathbb{R}^n$ for $n \geq 2$ be endowed with the standard Euclidean inner product $\langle \cdot \mid \cdot \rangle$. The associated norm is given by $\|x\| := \sqrt{\langle x \mid x \rangle}$. Let δ be any fixed positive real number and $f : \mathbb{R}^n \to \mathbb{R}^n$ any map such that $\|x-y\| = \delta \Rightarrow \|f(x) - f(y)\| = \delta$; then f is a Euclidean motion, i.e. $f \in \mathbb{R}^n \rtimes \mathsf{O}(n)$.*

Note that there are three obvious points which let the result of Beckman and Quarles in Euclidean space appear somewhat stronger than the theorem of Alexandrov in Minkowski space:

1. The conclusion of Theorem 3 holds for *any* $\delta \in \mathbb{R}_+$, whereas Alexandrov's theorem singles out lightlike distances.
2. In Theorem 3, $n = 2$ is not excluded.
3. In Theorem 3, f is not required to be a bijection, so that we did not assume the existence of an inverse map f^{-1}. Correspondingly, there is no assumption that f^{-1} also preserves the distance δ.

[16] In fact, Beckman and Quarles proved the conclusion of Theorem 3 under slightly weaker hypotheses: They allowed the map f to be "many-valued", that is, to be a map $f : \mathbb{R}^n \to \mathcal{S}^n$, where $\mathcal{S}^n$ is the set of non-empty subsets of $\mathbb{R}^n$, such that $\|x-y\| = \delta \Rightarrow \|x'-y'\| = \delta$ for any $x' \in f(x)$ and any $y' \in f(y)$. However, given the statement of Theorem 3, it is immediate that such "many-valued maps" must necessarily be single-valued. To see this, assume that $x_* \in \mathbb{R}^n$ has the two image points y_1, y_2 and define $h_i : \mathbb{R}^n \to \mathbb{R}^n$ for $i = 1, 2$ such that $h_1(x) = h_2(x) \in f(x)$ for all $x \neq x_*$ and $h_i(x_*) = y_i$. Then, according to Theorem 3, h_i must both be Euclidean motions. Since they are continuous and coincide for all $x \neq x_*$, they must also coincide at x_*.

5.3 Einstein Synchronization

We start by characterizing those cases in which a strict inverted Cauchy-Schwarz inequality holds:

Lemma 1. *Let V be of dimension $n > 2$ and $v \in V$ be some non-zero vector. The strict inverted Cauchy-Schwarz inequality,*

$$v^2 w^2 < (v \cdot w)^2 \,, \tag{96}$$

holds for all $w \in V$ linearly independent of v iff v is timelike.

Proof. Obviously v cannot be spacelike, for then we would violate (96) with any spacelike w. If v is lightlike then w violates (96) iff it is in the set $v^\perp - \mathrm{span}\{v\}$, which is non-empty iff $n > 2$. Hence v cannot be lightlike if $n > 2$. If v is timelike we decompose $w = av + w'$ with $w' \in v^\perp$ so that $w'^2 \leq 0$, with equality iff v and w are linearly dependent. Hence

$$(v \cdot w)^2 - v^2 w^2 = -v^2\, w'^2 \geq 0 \,, \tag{97}$$

with equality iff v and w are linearly dependent. □

The next Lemma deals with the intersection of a causal line with a light cone, a situation depicted in Fig. 2.

Lemma 2. *Let $\mathcal{L}_p$ be the light-doublecone with vertex p and $\ell := \{r + \lambda v \mid r \in \mathbb{R}\}$ be a non-spacelike line, i.e. $v^2 \geq 0$, through $r \notin \mathcal{L}_p$. If v is timelike $\ell \cap \mathcal{L}_p$ consists of two points. If v is lightlike this intersection consists of one point if $p - r \notin v^\perp$ and is empty if $p - r \in v^\perp$. Note that the latter two statements are independent of the choice of $r \in \ell$–as they must be–, i.e. are invariant under $r \mapsto r' := r + \sigma v$, where $\sigma \in \mathbb{R}$.*

Proof. We have $r + \lambda v \in \mathcal{L}_p$ iff

$$(r + \lambda v - p)^2 = 0 \iff \lambda^2 v^2 + 2\lambda v \cdot (r - p) + (r - p)^2 = 0 \,. \tag{98}$$

For v timelike we have $v^2 > 0$ and (98) has two solutions

$$\lambda_{1,2} = \frac{1}{v^2} \left\{ -v \cdot (r - p) \pm \sqrt{\big(v \cdot (r - p)\big)^2 - v^2 (r - p)^2} \right\} . \tag{99}$$

Indeed, since $r \notin \mathcal{L}_p$, the vectors v and $r - p$ cannot be linearly dependent so that Lemma 1 implies the positivity of the expression under the square root. If v is lightlike (98) becomes a linear equation which is has one solution if $v \cdot (r - p) \neq 0$ and no solution if $v \cdot (r - p) = 0$ [note that $(r - p)^2 \neq 0$ since $q \notin \mathcal{L}_p$ by hypothesis]. □

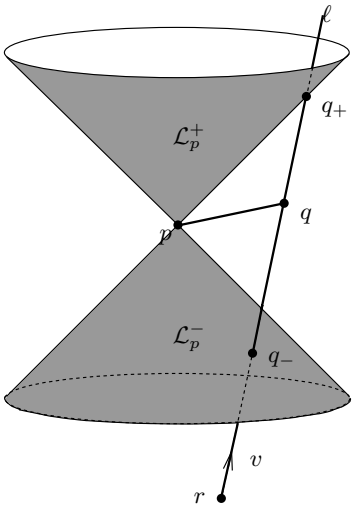


Fig. 2. A timelike line $\ell = \{r + \lambda v \mid \lambda \in \mathbb{R}\}$ intersects the light-cone with vertex $p \notin \ell$ in two points: q_+, its intersection with the future light-cone and q_-, its intersection with past the light cone. q is a point inbetween q_+ and q_-

Proposition 4. *Let ℓ and $\mathcal{L}_p$ as in Lemma 2 with v timelike. Let q_+ and q_- be the two intersection points of ℓ with $\mathcal{L}_p$ and $q \in \ell$ a point between them. Then*

$$\|q - p\|_g^2 = \|q_+ - q\|_g \, \|q - q_-\|_g \,. \tag{100}$$

Moreover, $\|q_+ - q\|_g = \|q - q_-\|_g$ iff $p - q$ is perpendicular to v.

Proof. The vectors $(q_+ - p) = (q - p) + (q_+ - q)$ and $(q_- - p) = (q - p) + (q_- - q)$ are lightlike, which gives (note that $q - p$ is spacelike):

$$\|q - p\|_g^2 \;=\; -(q - p)^2 \;=\; (q_+ - q)^2 + 2(q - p) \cdot (q_+ - q)\,, \tag{101a}$$

$$\|q - p\|_g^2 \;=\; -(q - p)^2 \;=\; (q_- - q)^2 + 2(q - p) \cdot (q_- - q)\,. \tag{101b}$$

Since $q_+ - q$ and $q - q_-$ are parallel we have $q_+ - q = \lambda(q - q_-)$ with $\lambda \in \mathbb{R}_+$ so that $(q_+ - q)^2 = \lambda \|q_+ - q\|_g \|q - q_-\|_g$ and $\lambda(q_- - q)^2 = \|q_+ - q\|_g \|q - q_-\|_g$. Now, multiplying (101b) with λ and adding this to (101a) immediately yields

$$(1 + \lambda)\, \|q - p\|_g^2 = (1 + \lambda)\; \|q_+ - q\|_g \|q - q_-\|_g \,. \tag{102}$$

Since $1 + \lambda \neq 0$ this implies (100). Finally, since $q_+ - q$ and $q_- - q$ are antiparallel, $\|q_+ - q\|_g = \|q_- - q\|_g$ iff $(q_+ - q) = -(q_- - q)$. Equations (101) now show that this is the case iff $(q - p) \cdot (q_\pm - q) = 0$, i.e. iff $(q - p) \cdot v = 0$. Hence we have shown

$$\|q_+ - q\|_g = \|q - q_-\|_g \iff (q - p) \cdot v = 0\,. \tag{103}$$

In other words, q is the midpoint of the segment $\overline{q_+ q_-}$ iff the line through p and q is perpendicular (wrt. g) to ℓ. □

The somewhat surprising feature of the first statement of this proposition is that (100) holds for *any* point of the segment $\overline{q_+q_-}$, not just the midpoint, as it would have to be the case for the corresponding statement in Euclidean geometry.

The second statement of Proposition 4 gives a convenient geometric characterization of Einstein-simultaneity. Recall that an event q on a timelike line ℓ (representing an inertial observer) is defined to be Einstein-simultaneous with an event p in spacetime iff q bisects the segment $\overline{q_+q_-}$ between the intersection points q_+, q_- of ℓ with the double-lightcone at p. Hence Proposition 4 implies

Corollary 1. *Einstein simultaneity with respect to a timelike line ℓ is an equivalence relation on spacetime, the equivalence classes of which are the spacelike hyperplanes orthogonal (wrt. g) to ℓ.*

The first statement simply follows from the fact that the family of parallel hyperplanes orthogonal to ℓ form a partition (cf. Sect. A.1) of spacetime.

From now on we shall use the terms "timelike line" and "inertial observer" synonymously. Note that Einstein simultaneity is only defined relative to an inertial observer. Given two inertial observers,

$$\ell \;=\; \{r + \lambda v \mid \lambda \in \mathbb{R}\} \qquad \text{first observer}\,, \tag{104a}$$

$$\ell' \;=\; \{r' + \lambda' v' \mid \lambda' \in \mathbb{R}\} \qquad \text{second observer}\,, \tag{104b}$$

we call the corresponding Einstein-simultaneity relations *ℓ-simultaneity* and *ℓ'-simultaneity*. Obviously they coincide iff ℓ and ℓ' are parallel (v and v' are linearly dependent). In this case $q' \in \ell'$ is ℓ-simultaneous to $q \in \ell$ iff $q \in \ell$ is ℓ'-simultaneous to $q' \in \ell'$. If ℓ and $\ell'$are not parallel (skew or intersecting in one point) it is generally not true that if $q' \in \ell'$ is ℓ-simultaneous to $q \in \ell$ then $q \in \ell$ is also ℓ'-simultaneous to $q' \in \ell'$. In fact, we have

Proposition 5. *Let ℓ and ℓ' two non-parallel timelike likes. There exists a unique pair $(q, q') \in \ell \times \ell'$ so that q' is ℓ-simultaneous to q and q is ℓ' simultaneous to q'.*

Proof. We parameterize ℓ and ℓ' as in (104). The two conditions for q' being ℓ-simultaneous to q and q being ℓ'-simultaneous to q' are $(q-q')\cdot v = 0 = (q-q')\cdot v'$. Writing $q = r + \lambda v$ and $q' = r' + \lambda' v'$ this takes the form of the following matrix equation for the two unknowns λ and λ':

$$\begin{pmatrix} v^2 & -v\cdot v' \\ v\cdot v' & -v'^2 \end{pmatrix} \begin{pmatrix} \lambda \\ \lambda' \end{pmatrix} = \begin{pmatrix} (r'-r)\cdot v \\ (r'-r)\cdot v' \end{pmatrix}. \tag{105}$$

This has a unique solution pair (λ, λ'), since for linearly independent timelike vectors v and v' Lemma 1 implies $(v\cdot v')^2 - v^2 v'^2 > 0$. Note that if ℓ and ℓ' intersect $q = q' =$ intersection point. □

Clearly, Einstein-simultaneity is conventional and physics proper should not depend on it. For example, the fringe-shift in the Michelson-Morley experiment is independent of how we choose to synchronize clocks. In fact, it does not

even make use of any clock. So what is the general definition of a "simultaneity structure"? It seems obvious that it should be a relation on spacetime that is at least symmetric (each event should be simultaneous to itself). Going from one-way simultaneity to the mutual synchronization of two clocks, one might like to also require reflexivity (if p is simultaneous to q then q is simultaneous to p), though this is not strictly required in order to one-way synchronize each clock in a set of clocks with one preferred "master clock", which is sufficient for many applications.

Moreover, if we like to speak of the mutual simultaneity of sets of more than two events we need an equivalence relation on spacetime. The equivalence relation should be such that each inertial observer intersect each equivalence class precisely once. Let us call such a simultaneity structure "admissible". Clearly there are zillions of such structures: just partition spacetime into any set of appropriate[17] spacelike hypersurfaces (there are more possibilities at this point, like families of forward or backward lightcones). An *absolute* admissible simultaneity structure would be one which is invariant (cf. Sect. A.1) under the automorphism group of spacetime. We have

Proposition 6. *There exits precisely one admissible simultaneity structure which is invariant under the inhomogeneous proper orthochronous Galilei group and none that is invariant under the inhomogeneous proper orthochronous Lorentz group.*

Proof. See [24]. □

There is a group-theoretic reason that highlights this existential difference:

Proposition 7. *Let G be a group with transitive action on a set S. Let* $\mathrm{Stab}(p) \subset G$ *be the stabilizer subgroup for $p \in S$ (due to transitivity all stabilizer subgroups are conjugate). Then S admits a G-invariant equivalence relation $R \subset S \times S$ iff* $\mathrm{Stab}(p)$ *is* not *maximal, that is, iff* $\mathrm{Stab}(p)$ *is properly contained in a proper subgroup H of G:* $\mathrm{Stab}(p) \subsetneq H \subsetneq G$.

Proof. See Theorem 1.12 in [32]. □

Regarding the action of the inhomogeneous Galilei and Lorentz groups on spacetime their stabilizers are the corresponding homogeneous groups. As already discussed at the end of Sect. 4.1, the homogeneous Lorentz group is maximal in the inhomogeneous one, whereas the homogeneous Galilei group is not maximal in the inhomogeneous one. This, according to Proposition 7, is the group theoretic origin of the absence of any invariant simultaneity structure in the Lorentzian case.

[17] For example, the hypersurfaces should not be asymptotically hyperboloidal, for then a constantly accelerated observer would not intersect all of them.

5.4 The Lattice Structure of Causally and Chronologically Complete Sets

Here we wish to briefly discuss another important structure associated with causality relations in Minkowski space, which plays a fundamental rôle in modern Quantum Field Theory (see e.g. [27]). Let S_1 and S_2 be subsets of $\mathbb{M}^n$. We say that S_1 and S_2 are *causally disjoint* or *spacelike separated* iff $p_1 - p_2$ is spacelike, i.e. $(p_1 - p_2)^2 < 0$, for any $p_1 \in S_1$ and $p_2 \in S_2$. Note that because a point is not spacelike separated from itself, causally disjoint sets are necessarily disjoint in the ordinary set-theoretic sense – the converse being of course not true.

For any subset $S \subseteq \mathbb{M}^n$ we denote by S' the largest subset of $\mathbb{M}^n$ which is causally disjoint to S. The set S' is called the *causal complement* of S. The procedure of taking the causal complement can be iterated and we set $S'' := (S')'$ etc. S'' is called the *causal completion* of S. It also follows straight from the definition that $S_1 \subseteq S_2$ implies $S_1' \supseteq S_2'$ and also $S'' \supseteq S$. If $S'' = S$ we call S *causally complete*. We note that the causal complement S' of any given S is automatically causally complete. Indeed, from $S'' \supseteq S$ we obtain $(S')'' \subseteq S'$, but the first inclusion applied to S' instead of S leads to $(S')'' \supseteq S'$, showing $(S')'' = S'$. Note also that for any subset S its causal completion, S'', is the smallest causally complete subset containing S, for if $S \subseteq K \subseteq S''$ with $K'' = K$, we derive from the first inclusion by taking $''$ that $S'' \subseteq K$, so that the second inclusion yields $K = S''$. Trivial examples of causally complete subsets of $\mathbb{M}^n$ are the empty set, single points, and the total set $\mathbb{M}^n$. Others are the open diamond-shaped regions (95) as well as their closed counterparts:

$$\bar{U}(p,q) := (\bar{\mathcal{C}}_p^+ \cap \bar{\mathcal{C}}_q^-) \cup (\bar{\mathcal{C}}_q^+ \cap \bar{\mathcal{C}}_p^-)\,. \tag{106}$$

We now focus attention to the set $\mathrm{Caus}(\mathbb{M}^n)$ of causally complete subsets of $\mathbb{M}^n$, including the empty set, $\emptyset$, and the total set, $\mathbb{M}^n$, which are mutually causally complementary. It is partially ordered by ordinary set-theoretic inclusion ($\subseteq$) (cf. Sect. A.1) and carries the "dashing operation" ($'$) of taking the causal complement. Moreover, on $\mathrm{Caus}(\mathbb{M}^n)$ we can define the operations of "meet" and "join", denoted by $\wedge$ and $\vee$ respectively, as follows: Let $S_i \in \mathrm{Caus}(\mathbb{M}^n)$ where $i = 1, 2$, then $S_1 \wedge S_2$ is the largest causally complete subset in the intersection $S_1 \cap S_2$ and $S_1 \vee S_2$ is the smallest causally complete set containing the union $S_1 \cup S_2$.

The operations of $\wedge$ and $\vee$ can be characterized in terms of the ordinary set-theoretic intersection $\cap$ together with the dashing-operation. To see this, consider two causally complete sets, S_i where $i = 1, 2$, and note that the set of points that are spacelike separated from S_1 and S_2 are obviously given by $S_1' \cap S_2'$, but also by $(S_1 \cup S_2)'$, so that

$$S_1' \cap S_2' \;=\; (S_1 \cup S_2)'\,, \tag{107a}$$

$$S_1 \cap S_2 \;=\; (S_1' \cup S_2')'\,. \tag{107b}$$

Here (107a) and (107b) are equivalent since any $S_i \in \mathrm{Caus}(\mathbb{M}^n)$ can be written as $S_i = P_i'$, namely $P_i = S_i'$. If S_i runs through all sets in $\mathrm{Caus}(\mathbb{M}^n)$ so does P_i.

Hence any equation that holds generally for all $S_i \in \mathrm{Caus}(\mathbb{M}^n)$ remains valid if the S_i are replaced by S_i'.

Equation (107b) immediately shows that $S_1 \cap S_2$ is causally complete (since it is the $'$ of something). Taking the causal complement of (107a) we obtain the desired relation for $S_1 \vee S_2 := (S_1 \cup S_2)''$. Together we have

$$S_1 \wedge S_2 \;=\; S_1 \cap S_2\,, \tag{108a}$$

$$S_1 \vee S_2 \;=\; (S_1' \cap S_2')'\,. \tag{108b}$$

From these we immediately derive

$$(S_1 \wedge S_2)' \;=\; S_1' \vee S_2'\,, \tag{109a}$$

$$(S_1 \vee S_2)' \;=\; S_1' \wedge S_2'\,. \tag{109b}$$

All what we have said so far for the set $\mathrm{Caus}(\mathbb{M}^n)$ could be repeated verbatim for the set $\mathrm{Chron}(\mathbb{M}^n)$ of *chronologically complete* subsets. We say that S_1 and S_2 are *chronologically disjoint* or *non-timelike separated*, iff $S_1 \cap S_2 = \emptyset$ and $(p_1 - p_2)^2 \leq 0$ for any $p_1 \in S_1$ and $p_2 \in S_2$. S', the *chronological complement* of S, is now the largest subset of $\mathbb{M}^n$ which is chronologically disjoint to S. The only difference between the causal and the chronological complement of S is that the latter now contains lightlike separated points, which are not contained in S. A set S is *chronologically complete* iff $S = S''$, where the dashing now denotes the operation of taking the chronological complement. Again, for any set S the set S' is automatically chronologically complete and S'' is the smallest chronologically complete subset containing S. Single points are chronologically complete subsets and every chronologically complete subset is the join of its points. All the formal properties regarding $'$, $\wedge$, and $\vee$ stated hitherto for $\mathrm{Caus}(\mathbb{M}^n)$ are the same for $\mathrm{Chron}(\mathbb{M}^n)$.

One major difference between $\mathrm{Caus}(\mathbb{M}^n)$ and $\mathrm{Chron}(\mathbb{M}^n)$ is that the types of diamond-shaped sets they contain are different. For example, the closed ones, (106), are members of both. The open ones, (95), are contained in $\mathrm{Caus}(\mathbb{M}^n)$ but *not* in $\mathrm{Chron}(\mathbb{M}^n)$. Instead, $\mathrm{Chron}(\mathbb{M}^n)$, contains the closed diamonds whose 'equator'[18] have been removed. An essential structural difference between $\mathrm{Caus}(\mathbb{M}^n)$ and $\mathrm{Chron}(\mathbb{M}^n)$ will be stated below, after we have introduced the notion of a lattice to which we now turn.

To put all these formal properties into the right frame we recall the definition of a lattice. Let $(L, \leq)$ be a partially ordered set and a, b any two elements in L. Synonymously with $a \leq b$ we also write $b \geq a$ and say that a is smaller than b, b is bigger than a, or b majorizes a. We also write $a < b$ if $a \leq b$ and $a \neq b$. If, with respect to $\leq$, their greatest lower and least upper bound exist, they are denoted by $a \wedge b$–called the "meet of a and b"–and $a \vee b$–called the "join of a and b''–respectively. A partially ordered set for which the greatest lower and least upper bound exist for any pair a, b of elements from L is called a *lattice*.

[18] By "equator" we mean the $(n-2)$–sphere in which the forward and backward light-cones in (106) intersec. In the two-dimensional drawings the "equator" is represented by just two points marking the right and left corners of the diamond-shaped set.

We now list some of the most relevant additional structural elements lattices can have: A lattice is called *complete* if greatest lower and least upper bound exist for any subset $K \subseteq L$. If $K = L$ they are called 0 (the smallest element in the lattice) and 1 (the biggest element in the lattice) respectively. An atom in a lattice is an element a which majorizes only 0, i.e. $0 \leq a$, and if $0 \leq b \leq a$ then $b = 0$ or $b = a$. The lattice is called *atomic* if each of its elements different from 0 majorizes an atom. An atomic lattice is called *atomistic* if every element is the join of the atoms it majorizes. An element c is said to *cover* a if $a < c$ and if $a \leq b \leq c$ either $a = b$ or $b = c$. An atomic lattice is said to have the *covering property* if for every element b and every atom a for which $a \wedge b = 0$ the join $a \vee b$ covers b.

The subset $\{a, b, c\} \subseteq L$ is called a *distributive triple* if

$$a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c) \quad \text{and } (a, b, c) \text{ cyclically permuted}\,, \tag{110a}$$
$$a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c) \quad \text{and } (a, b, c) \text{ cyclically permuted}\,. \tag{110b}$$

Definition 5. *A lattice is called* distributive *or* Boolean *if every triple* $\{a, b, c\}$ *is distributive. It is called* modular *if every triple* $\{a, b, c\}$ *with* $a \leq b$ *is distributive.*

It is straightforward to check from (110) that modularity is equivalent to a single condition as follows:

$$\text{modularity} \Leftrightarrow a \vee (b \wedge c) = b \wedge (a \vee c) \quad \text{for all } a, b, c \in L \text{ s.t. } a \leq b. \tag{111}$$

If in a lattice with smallest element 0 and greatest element 1 a map $L \to L$, $a \mapsto a'$, exist such that

$$a'' := (a')' = a\,, \tag{112a}$$
$$a \leq b \Rightarrow b' \leq a'\,, \tag{112b}$$
$$a \wedge a' = 0\,, \quad a \vee a' = 1\,, \tag{112c}$$

the lattice is called *orthocomplemented*. It follows that whenever the meet and join of a subset $\{a_i \mid i \in I\}$ (I is some index set) exist, one has De Morgan's laws:[19]

$$\big(\textstyle\bigwedge_{i\in I}\, a_i\big)' = \bigvee_{i\in I}\, a_i'\,, \tag{113a}$$
$$\big(\textstyle\bigvee_{i\in I}\, a_i\big)' = \bigwedge_{i\in I}\, a_i'\,. \tag{113b}$$

For orthocomplemented lattices there is a still weaker version of distributivity than modularity, which turns out to be physically relevant in various contexts:

Definition 6. *An orthocomplemented lattice is called* orthomodular *if every triple* $\{a, b, c\}$ *with* $a \leq b$ *and* $c \leq b'$ *is distributive.*

[19] From these laws it also appears that the definition (112c) is redundant, as each of its two statements follows from the other, due to $0' = 1$.

From (111) and using that $b \wedge c = 0$ for $b \leq c'$ one sees that this is equivalent to the single condition (renaming c to c'):

$$\text{orthomod.} \;\Leftrightarrow\; a = b \wedge (a \vee c') \quad \text{for all } a, b, c \in L \text{ s.t. } a \leq b \leq c\,, \tag{114a}$$

$$\Leftrightarrow\; a = b \vee (a \wedge c') \quad \text{for all } a, b, c \in L \text{ s.t. } a \geq b \geq c\,, \tag{114b}$$

where the second line follows from the first by taking its orthocomplement and renaming a', b', c to a, b, c'. It turns out that these conditions can still be simplified by making them independent of c. In fact, (114) are equivalent to

$$\text{orthomod.} \;\Leftrightarrow\; a = b \wedge (a \vee b') \quad \text{for all } a, b \in L \text{ s.t. } a \leq b\,, \tag{115a}$$

$$\Leftrightarrow\; a = b \vee (a \wedge b') \quad \text{for all } a, b \in L \text{ s.t. } a \geq b\,. \tag{115b}$$

It is obvious that (114) implies (115) (set $c = b$). But the converse is also true. To see this, take e.g. (115b) and choose any $c \leq b$. Then $c' \geq b'$, $a \geq b$ (by hypothesis), and $a \geq a \wedge c'$ (trivially), so that $a \geq b \vee (a \wedge c')$. Hence $a \geq b \vee (a \wedge c') \geq b \vee (a \wedge b') = a$, which proves (114b).

Complete orthomodular atomic lattices are automatically atomistic. Indeed, let b be the join of all atoms majorized by $a \neq 0$. Assume $a \neq b$ so that necessarily $b < a$, then (115b) implies $a \wedge b' \neq 0$. Then there exists an atom c majorized by $a \wedge b'$. This implies $c \leq a$ and $c \leq b'$, hence also $c \not\leq b$. But this is a contradiction, since b is by definition the join of all atoms majorized by a.

Finally we mention the notion of *compatibility* or *commutativity*, which is a symmetric, reflexive, but generally not transitive relation R on an orthomodular lattice (cf. Sect. A.1). We write $a \natural b$ for $(a, b) \in R$ and define:

$$a \natural b \;\Leftrightarrow\; a = (a \wedge b) \vee (a \wedge b')\,, \tag{116a}$$

$$\Leftrightarrow\; b = (b \wedge a) \vee (b \wedge a')\,. \tag{116b}$$

The equivalence of these two lines, which shows that the relation of being compatible is indeed symmetric, can be demonstrated using orthomodularity as follows: Suppose (116a) holds; then $b \wedge a' = b \wedge (b' \vee a') \wedge (b \vee a') = b \wedge (b' \vee a')$, where we used the orthocomplement of (116a) to replace a' in the first expression and the trivial identity $b \wedge (b \vee a') = b$ in the second step. Now, applying (115b) to $b \geq a \wedge b$ we get $b = (b \wedge a) \vee [b \wedge (b' \vee a')] = (b \wedge a) \vee (b \wedge a')$, i.e. (116b). The converse, (116*b*) $\Rightarrow$ (116*a*), is of course entirely analogous.

From (116) a few things are immediate: $a \natural b$ is equivalent to $a \natural b'$, $a \natural b$ is implied by $a \leq b$ or $a \leq b'$, and the elements 0 and 1 are compatible with all elements in the lattice. The *center* of a lattice is the set of elements which are compatible with all elements in the lattice. In fact, the center is a Boolean sublattice. If the center contains no other elements than 0 and 1 the lattice is said to be *irreducible*. The other extreme is a Boolean lattice, which is identical to its own center. Indeed, if (a, b, b') is a distributive triple, one has $a = a \wedge 1 = a \wedge (b \vee b') = (a \wedge b) \vee (a \wedge b') \Rightarrow$ (116*a*).

After these digression into elementary notions of lattice theory we come back to our examples of the sets $\mathrm{Caus}(\mathbb{M}^n)$ $\mathrm{Chron}(\mathbb{M}^n)$. Our statements above amount

to saying that they are complete, atomistic, and orthocomplemented lattices. The partial order relation $\leq$ is given by $\subseteq$, and the extreme elements 0 and 1 correspond to the empty set $\emptyset$ and the total set $\mathbb{M}^n$, the points of which are the atoms. Neither the covering property nor modularity is shared by any of the two lattices, as can be checked by way of elementary counterexamples.[20] In particular, neither of them is Boolean. However, in [15] it was shown that $\mathrm{Chron}(\mathbb{M}^n)$ is orthomodular; see also [13], which deals with more general spacetimes. In contrast, $\mathrm{Caus}(\mathbb{M}^n)$ is definitely *not* orthomodular, as is e.g. seen by the counterexample given in Fig. 3.[21] It is also not difficult to prove that $\mathrm{Chron}(\mathbb{M}^n)$ is irreducible.[22]

It is well known that the lattices of propositions for classical systems are Boolean, whereas those for quantum systems are merely orthomodular. In classical physics the elements of the lattice are measurable subsets of phase space, with $\leq$ being ordinary set-theoretic inclusion $\subseteq$, and $\wedge$ and $\vee$ being ordinary set-theoretic intersection $\cap$ and union $\cup$ respectively. The orthocomplement is the ordinary set-theoretic complement. In Quantum Mechanics the elements of the lattice are the closed subspaces of Hilbert space, with $\leq$ being again ordinary inclusion, $\wedge$ ordinary intersection, and $\vee$ is given by $a \vee b := \overline{\mathrm{span}\{a, b\}}$. The orthocomplement of a closed subset is the orthogonal complement in Hilbert space. For more information see [33] and [5].

One of the main questions in the foundations of Quantum Mechanics is whether one could understand (derive) the usage of Hilbert spaces and complex numbers from somehow more fundamental principles. Even though it is not a priori clear what ones measure of fundamentality should be at this point, an interesting line of attack consists in deriving the mentioned structures from the properties of the lattice of propositions (Quantum Logic). It can be shown that a lattice that is complete, atomic, irreducible, orthomodular, and that satisfies the covering property, is isomorphic to the lattice of closed subspaces of a linear space with Hermitean inner product. The complex numbers are selected if

[20] An immediate counterexample for the covering property is this: Take two timelike separated points (i.e. atoms) p and q. Then $\{p\} \wedge \{q\} = \emptyset$ whereas $\{p\} \vee \{q\}$ is given by the closed diamond (106). Note that this is true in $\mathrm{Caus}(\mathbb{M}^n)$ *and* $\mathrm{Chron}(\mathbb{M}^n)$. But, clearly, $\{p\} \vee \{q\}$ does not cover either $\{p\}$ or $\{q\}$.

[21] Regarding this point, there are some conflicting statements in the literature. The first edition of [27] states orthomodularity of $\mathrm{Chron}(\mathbb{M}^n)$ in Proposition 4.1.3, which is removed in the second edition without further comment. The proof offered in the first edition uses (115a) as definition of orthomodularity, writing K_1 for a and K_2 for b. The crucial step is the claim that any spacetime event in the set $K_2 \wedge (K_1 \vee K_2')$ lies in K_2 and that any causal line through it must intersect either K_1 or K_2'. The last statement is, however, not correct since the join of two sets (here K_1 and K_2') is generally larger than the domain of dependence of their ordinary set-theoretic union; compare Fig. 3. : (Generally, the domain of dependence of a subset S of spacetime M is the largest subset $D(S) \subseteq M$ such that any inextensible causal curve that intersects $D(S)$ also intersects S.)

[22] In general spacetimes M, the failure of irreducibility of $\mathrm{Chron}(M)$ is directly related to the existence of closed timelike curves; see [13].

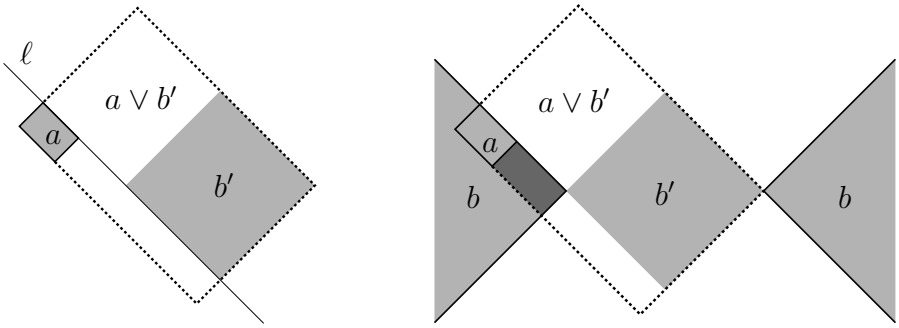


Fig. 3. The two figures show that Caus($\mathbb{M}^n$) is not orthomodular. The first thing to note is that Caus($\mathbb{M}^n$) contains open (95) as well as closed (106) diamond sets. In the left picture we consider the join of a small closed diamond a with a large open diamond b'. (Closed sets are indicated by a solid boundary line.) Their edges are aligned along the lightlike line ℓ. Even though these regions are causally disjoint, their causal completion is much larger than their union and given by the open (for $n > 2$) enveloping diamond $a \vee b'$ framed by the dashed line. (This also shows that the join of two regions can be larger than the domain of dependence of their union; compare footnote 21.) . Next we consider the situation depicted on the right side. The closed double-wedge region b contains the small closed diamond a. The causal complement b' of b is the open diamond in the middle. $a \vee b'$ is, according to the first picture, given by the large open diamond enclosed by the dashed line. The intersection of $a \vee b'$ with b is strictly larger than a, the difference being the dark-shaded region in the left wedge of b below a. Hence $a \neq b \wedge (a \vee b')$, in contradiction to (115a)

additional technical assumptions are added. For the precise statements of these reconstruction theorems see [5].

It is now interesting to note that, on a formal level, there is a similar transition in going from Galilei invariant to Lorentz invariant causality relations. In fact, in Galilean spacetime one can also define a chronological complement: Two points are chronologically related if they are connected by a worldline of finite speed and, accordingly, two subsets in spacetime are chronologically disjoint if no point in one set is chronologically related to a point of the other. For example, the chronological complement of a point p are all points simultaneous to, but different from, p. More general, it is not hard to see that the chronologically complete sets are just the subsets of some $t = \text{const.}$ hypersurface. The lattice of chronologically complete sets is then the continuous disjoint union of sublattices, each of which is isomorphic to the Boolean lattice of subsets in $\mathbb{R}^3$. For details see [14].

As we have seen above, Chron($\mathbb{M}^n$) is complete, atomic, irreducible, and orthomodular. The main difference to the lattice of propositions in Quantum Mechanics, as regards the formal aspects discussed here, is that Chron($\mathbb{M}^n$) does *not* satisfy the covering property. Otherwise the formal similarities are intriguing and it is tempting to ask whether there is a deeper meaning to this. In this respect it would be interesting to know whether one could give a lattice-

theoretic characterization for Chron(M) (M some fixed spacetime), comparable to the characterization of the lattices of closed subspaces in Hilbert space alluded to above. Even for $M = \mathbb{M}^n$ such a characterization seems, as far as I am aware, not to be known.

5.5 Rigid Motion

As is well known, the notion of a rigid body, which proves so useful in Newtonian mechanics, is incompatible with the existence of a universal finite upper bound for all signal velocities [36]. As a result, the notion of a perfectly rigid body does not exist within the framework of SR. However, the notion of a *rigid motion* does exist. Intuitively speaking, a body moves rigidly if, locally, the relative spatial distances of its material constituents are unchanging.

The motion of an extended body is described by a normalized timelike vector field $u : \Omega \to \mathbb{R}^n$, where Ω is an open subset of Minkowski space, consisting of the events where the material body in question "exists". We write $g(u,u) = u \cdot u = u^2$ for the Minkowskian scalar product. Being normalized now means that $u^2 = c^2$ (we do *not* choose units such that $c = 1$). The Lie derivative with respect to u is denoted by L_u.

For each material part of the body in motion its local rest space at the event $p \in \Omega$ can be identified with the hyperplane through p orthogonal to u_p:

$$H_p := p + u_p^{\perp} \,. \tag{117}$$

$u_p^{\perp}$ carries a Euclidean inner product, h_p, given by the restriction of $-g$ to $u_p^{\perp}$. Generally we can write

$$h = c^{-2}\, u^\flat \otimes u^\flat - g \,, \tag{118}$$

where $u^\flat = g^\downarrow(u) := g(u,\cdot)$ is the one-form associated ('index-lowered', cf. Sect. A.5) to u. Following [12] the precise definition of "rigid motion" can now be given as follows:

Definition 7 (Born 1909). *Let u be a normalized timelike vector field u. The motion described by its flow is* rigid *if*

$$L_u h = 0 \,. \tag{119}$$

Note that, in contrast to the Killing equations $L_u g = 0$, these equations are non linear due to the dependence of h upon u.

We write $\Pi_h := \mathrm{id} - c^{-2}\, u \otimes u^\flat \in \mathrm{End}(\mathbb{R}^n)$ for the tensor field over spacetime that pointwise projects vectors perpendicular to u. It acts on one forms α via $\Pi_h(\alpha) := \alpha \circ \Pi_h$ and accordingly on all tensors. The so extended projection map will still be denoted by Π_h. Then we e.g. have

$$h = -\Pi_h g := -g(\Pi_h \cdot, \Pi_h \cdot) \,. \tag{120}$$

It is not difficult to derive the following two equations:[23]

$$L_{fu}h = fL_u h\,, \tag{121}$$
$$L_u h = -L_u(\Pi_h g) = -\Pi_h(L_u g)\,, \tag{122}$$

where f is any differentiable real-valued function on Ω.

Equation (121) shows that the normalized vector field u satisfies (119) iff any rescaling fu with a nowhere vanishing function f does. Hence the normalization condition for u in (119) is really irrelevant. It is the geometry in spacetime of the flow lines and not their parameterization which decide on whether motions (all, i.e. for any parameterization, or none) along them are rigid. This has be the case because, generally speaking, there is no distinguished family of sections (hypersurfaces) across the bundle of flow lines that would represent "the body in space", i.e. mutually simultaneous locations of the body's points. Distinguished cases are those exceptional ones in which u is hypersurface orthogonal. Then the intersection of u's flow lines with the orthogonal hypersurfaces consist of mutually *Einstein synchronous* locations of the points of the body. An example is discussed below.

Equation (122) shows that the rigidity condition is equivalent to the "spatially" projected Killing equation. We call the flow of the timelike normalized vector field u a *Killing motion* (i.e. a spacetime isometry) if there is a Killing field K such that $u = cK/\sqrt{K^2}$. Equation (122) immediately implies that Killing motions are rigid. What about the converse? Are there rigid motions that are not Killing? This turns out to be a difficult question. Its answer in Minkowski space is: "yes, many, but not as many as naively expected."

Before we explain this, let us give an illustrative example for a Killing motion, namely that generated by the boost Killing-field in Minkowski space. We suppress all but one spatial directions and consider boosts in x direction in two-dimensional Minkowski space (coordinates ct and x; metric $ds^2 = c^2dt^2 - dx^2$). The Killing field is[24]

$$K = x\,\partial_{ct} + ct\,\partial_x\,, \tag{123}$$

which is timelike in the region $|x| > |ct|$. We focus on the "right wedge" $x > |ct|$, which is now our region Ω. Consider a rod of length ℓ which at $t = 0$ is represented by the interval $x \in (r, r+\ell)$, where $r > 0$. The flow of the normalized field $u = cK/\sqrt{K^2}$ is

[23] Equation (122) simply follows from $L_u\Pi_h = -c^{-2}u \otimes L_u u^\flat$, so that $g((L_u\Pi_h)X, \Pi_h Y) = 0$ for all X, Y. In fact, $L_u u^\flat = a^\flat$, where $a := \nabla_u u$ is the spacetime-acceleration. This follows from $L_u u^\flat(X) = L_u(g(u, X)) - g(u, L_u X) = g(\nabla_u u, X) + g(u, \nabla_u X - [u, X]) = g(a, X) - g(u, \nabla_X u) = g(a, X)$, where $g(u, u) =$ const. was used in the last step.

[24] Here we adopt the standard notation from differential geometry, where $\partial_\mu := \partial/\partial x^\mu$ denote the vector fields naturally defined by the coordinates $\{x^\mu\}_{\mu=0\cdots n-1}$. Pointwise the dual basis to $\{\partial_\mu\}_{\mu=0\cdots n-1}$ is $\{dx^\mu\}_{\mu=0\cdots n-1}$.

$$ct(\tau) \;=\; x_0\,\sinh\bigl(c\tau/x_0\bigr)\,, \tag{124a}$$

$$x(\tau) \;=\; x_0\,\cosh\bigl(c\tau/x_0\bigr)\,, \tag{124b}$$

where $x_0 = x(\tau = 0) \in (r, r+\ell)$ labels the elements of the rod at $\tau = 0$. We have $x^2 - c^2t^2 = x_0^2$, showing that the individual elements of the rod move on hyperbolae ('hyperbolic motion'). τ is the proper time along each orbit, normalized so that the rod lies on the x axis at $\tau = 0$.

The combination

$$\lambda := c\tau/x_0 \tag{125}$$

is just the flow parameter for K (123), sometimes referred to as "Killing time" (though it is dimensionless). From (124) we can solve for λ and τ as functions of ct and x:

$$\lambda \;=\; f(ct,x) \;:=\; \tanh^{-1}\bigl(ct/x\bigr)\,, \tag{126a}$$

$$\tau \;=\; \hat{f}(ct,x) \;:=\; \underbrace{\sqrt{(x/c)^2 - t^2}}_{x_0/c}\,\tanh^{-1}\bigl(ct/x\bigr)\,, \tag{126b}$$

from which we infer that the hypersurfaces of constant λ are hyperplanes which all intersect at the origin. Moreover, we also have $df = K^\flat/K^2$ (d is just the ordinary exterior differential) so that the hyperplanes of constant λ intersect all orbits of u (and K) orthogonally. Hence the hyperplanes of constant λ qualify as the equivalence classes of mutually Einstein-simultaneous events in the region $x > |ct|$ for a family of observers moving along the Killing orbits. This does not hold for the hypersurfaces of constant τ, which are curved.

The modulus of the spacetime-acceleration (which is the same as the modulus of the spatial acceleration measured in the local rest frame) of the material part of the rod labeled by x_0 is

$$\|a\|_g = c^2/x_0\,. \tag{127}$$

As an aside we generally infer from this that, given a timelike curve of local acceleration (modulus) α, infinitesimally nearby orthogonal hyperplanes intersect at a spatial distance c^2/α. This remark will become relevant in the discussion of part 2 of the Noether-Herglotz theorem given below.

In order to accelerate the rod to the uniform velocity v without deforming it, its material point labeled by x_0 has to accelerate for the eigentime (this follows from (124))

$$\tau = \frac{x_0}{c}\,\tanh^{-1}(v/c)\,, \tag{128}$$

which depends on x_0. In contrast, the Killing time is the same for all material points and just given by the final rapidity. In particular, judged from the local observers moving with the rod, a rigid acceleration requires accelerating the rod's trailing end harder but shorter than pulling its leading end.

In terms of the coordinates (λ, x_0), which are comoving with the flow of K, and (τ, x_0), which are comoving with the flow of u, we just have $K = \partial/\partial\lambda$ and $u = \partial/\partial\tau$ respectively. The spacetime metric g and the projected metric h in terms of these coordinates are:

$$h \;=\; dx_0^2\,, \tag{129a}$$

$$g \;=\; x_0^2\,d\lambda^2 - dx_0^2 = c^2\big(d\tau - (\tau/x_0)\,dx_0\big)^2 - dx_0^2\,. \tag{129b}$$

Note the simple form g takes in terms of x_0 and λ, which are also called the "Rindler coordinates" for the region $|x| > |ct|$ of Minkowski space. They are the analogs in Lorentzian geometry to polar coordinates (radius x_0, angle λ) in Euclidean geometry.

Let us now return to the general case. We decompose the derivative of the velocity one-form $u^\flat := g^\downarrow(u)$ as follows:

$$\nabla u^\flat = \theta + \omega + c^{-2}\,u^\flat \otimes a^\flat\,, \tag{130}$$

where θ and ω are the projected symmetrized and antisymmetrized derivatives respectively[25]

$$2\theta \;=\; \Pi_h(\nabla \vee u^\flat) \;=\; \nabla \vee u^\flat - c^{-2}\,u^\flat \vee a^\flat\,, \tag{131a}$$

$$2\omega \;=\; \Pi_h(\nabla \wedge u^\flat) \;=\; \nabla \wedge u^\flat - c^{-2}\,u^\flat \wedge a^\flat\,. \tag{131b}$$

The symmetric part, θ, is usually further decomposed into its traceless and pure trace part, called the *shear* and *expansion* of u respectively. The antisymmetric part ω is called the *vorticity* of u.

Now recall that the Lie derivative of g is just twice the symmetrized derivative:

$$L_u g = \nabla \vee u^\flat\,. \tag{132}$$

This implies in view of (119), (122), and (131a)

Proposition 8. *Let u be a normalized timelike vector field u. The motion described by its flow is rigid iff u is of vanishing shear and expansion, i.e. iff $\theta = 0$.*

Vector fields generating rigid motions are now classified according to whether or not they have a vanishing vorticity ω: if $\omega = 0$ the flow is called *irrotational*, otherwise *rotational*. The following theorem is due to Herglotz [29] and Noether [37]:

Theorem 4 (Noether & Herglotz, part 1). *A rotational rigid motion in Minkowski space must be a Killing motion.*

An example of such a rotational motion is given by the Killing field[26]

$$K = \partial_t + \kappa\,\partial_\varphi \tag{133}$$

[25] We denote the symmetrized and antisymmetrized tensor-product (not including the factor $1/n!$) by $\vee$ and $\wedge$ respectively and the symmetrized and antisymmetrized (covariant-) derivative by $\nabla\vee$ and $\nabla\wedge$. For example, $(u^\flat \wedge v^\flat)_{ab} = u_a v_b - u_b v_a$ and $(\nabla \vee u^\flat)_{ab} = \nabla_a u_b + \nabla_b u_a$. Note that $(\nabla \wedge u^\flat)$ is the same as the ordinary exterior differential $du^\flat$. Everything we say in the sequel applies to curved spacetimes if ∇ is read as covariant derivative with respect to the Levi-Civita connection.

[26] We now use standard cylindrical coordinates (z, ρ, φ), in terms of which $ds^2 = c^2dt^2 - dz^2 - d\rho^2 - \rho^2\,d\varphi^2$.

inside the region

$$\Omega = \{(t, z, \rho, \varphi) \mid \kappa\rho < c\}\,, \tag{134}$$

where K is timelike. This motion corresponds to a rigid rotation with constant angular velocity κ which, without loss of generality, we take to be positive. Using the comoving angular coordinate $\psi := \varphi - \kappa t$, the split (118) is now furnished by

$$u^\flat \;=\; c\sqrt{1-(\kappa\rho/c)^2}\left\{c\,dt - \frac{\kappa\rho/c}{1-(\kappa\rho/c)^2}\;\rho\,d\psi\right\}\,, \tag{135a}$$

$$h \;=\; dz^2 + d\rho^2 + \frac{\rho^2\,d\psi^2}{1-(\kappa\rho/c)^2}\,. \tag{135b}$$

The metric h is curved (cf. Lemma 3). But the rigidity condition (119) means that h, and hence its curvature, cannot change along the motion. Therefore, even though we can keep a body in uniform rigid rotational motion, we cannot put it into this state from rest by purely rigid motions, since this would imply a transition from a flat to a curved geometry of the body. This was first pointed out by Ehrenfest [17]. Below we will give a concise analytical expression of this fact (cf. equation (139)). All this is in contrast to the translational motion, as we will also see below.

The proof of Theorem 4 relies on arguments from differential geometry proper and is somewhat tricky. Here we present the essential steps, basically following [38] and [43] in a slightly modernized notation. Some straightforward calculational details will be skipped. The argument itself is best broken down into several lemmas.

At the heart of the proof lies the following general construction: Let M be the spacetime manifold with metric g and $\Omega \subset M$ the open region in which the normalized vector field u is defined. We take Ω to be simply connected. The orbits of u foliate Ω and hence define an equivalence relation on Ω given by $p \sim q$ iff p and q lie on the same orbit. The quotient space $\hat{\Omega} := \Omega/\!\sim$ is itself a manifold. Tensor fields on $\hat{\Omega}$ can be represented by (i.e. are in bijective correspondence to) tensor fields T on Ω which obey the two conditions:

$$\Pi_h\, T \;=\; T\,, \tag{136a}$$

$$L_u T \;=\; 0\,. \tag{136b}$$

Tensor fields satisfying (136a) are called *horizontal*, those satisfying both conditions (136) are called *projectable*. The $(n-1)$-dimensional metric tensor h, defined in (118), is an example of a projectable tensor if u generates a rigid motion, as assumed here. It turns $(\hat{\Omega}, h)$ into a $(n-1)$-dimensional Riemannian manifold. The covariant derivative $\hat{\nabla}$ with respect to the Levi-Civita connection of h is given by the following operation on projectable tensor fields:

$$\hat{\nabla} := \Pi_h \circ \nabla \tag{137}$$

i.e. by first taking the covariant derivative ∇ (Levi-Civita connection in (M, g)) in spacetime and then projecting the result horizontally. This results again in a projectable tensor, as a straightforward calculation shows.

The horizontal projection of the spacetime curvature tensor can now be related to the curvature tensor of $\hat{\Omega}$ (which is a projectable tensor field). Without proof we state

Lemma 3. *Let u generate a rigid motion in spacetime. Then the horizontal projection of the totally covariant (i.e. all indices down) curvature tensor R of (Ω, g) is related to the totally covariant curvature tensor $\hat{R}$ of $(\hat{\Omega}, h)$ by the following equation:*[27]

$$\Pi_h R = -\hat{R} - 3\,(\mathrm{id} - \Pi_\wedge)\omega \otimes \omega\,, \tag{138}$$

where $\Pi_\wedge$ is the total antisymmetrizer, which here projects tensors of rank four onto their totally antisymmetric part.

Formula (138) is true in any spacetime dimension n. Note that the projector $(\mathrm{id} - \Pi_\wedge)$ guarantees consistency with the first Bianchi identities for R and $\hat{R}$, which state that the total antisymmetrization in their last three slots vanish identically. This is consistent with (138) since for tensors of rank four with the symmetries of $\omega \otimes \omega$ the total antisymmetrization on tree slots is identical to $\Pi_\wedge$, the symmetrization on all four slots. The claim now simply follows from $\Pi_\wedge \circ (\mathrm{id} - \Pi_\wedge) = \Pi_\wedge - \Pi_\wedge = 0$.

We now restrict to spacetime dimensions of four or less, i.e. $n \leq 4$. In this case $\Pi_\wedge \circ \Pi_h = 0$ since Π_h makes the tensor effectively live over $n-1$ dimensions, and any totally antisymmetric four tensor in three or less dimensions must vanish. Applied to (138) this means that $\Pi_\wedge(\omega \otimes \omega) = 0$, for horizontality of ω implies $\omega \otimes \omega = \Pi_h(\omega \otimes \omega)$. Hence the right hand side of (138) just contains the pure tensor product $-3\,\omega \otimes \omega$.

Now, in our case $R = 0$ since (M, g) is flat Minkowski space. This has two interesting consequences: First, $(\hat{\Omega}, h)$ is curved iff the motion is rotational, as exemplified above. Second, since $\hat{R}$ is projectable, its Lie derivative with respect to u vanishes. Hence (138) implies $L_u\omega \otimes \omega + \omega \otimes L_u\omega = 0$, which is equivalent to[28]

$$L_u\omega = 0\,. \tag{139}$$

This says that the vorticity cannot change along a rigid motion in flat space. It is the precise expression for the remark above that you cannot rigidly *set* a disk into rotation. Note that it also provides the justification for the global classification of rigid motions into rotational and irrotational ones.

A sharp and useful criterion for whether a rigid motion is Killing or not is given by the following

Lemma 4. *Let u be a normalized timelike vector field on a region $\Omega \subseteq M$. The motion generated by u is Killing iff it is rigid and $a^\flat$ is exact on Ω.*

[27] $\hat{R}$ appears with a minus sign on the right hand side of (138) because the first index on the hatted curvature tensor is lowered with h rather than g. This induces a minus sign due to (118), i.e. as a result of our "mostly-minus"-convention for the signature of the spacetime metric.

[28] In more than four spacetime dimensions one only gets $(\mathrm{id} - \Pi_\wedge)(L_u\omega \otimes \omega + \omega \otimes L_u\omega) = 0$.

Proof. That the motion generated by u be Killing is equivalent to the existence of a positive function $f : \Omega \to \mathbb{R}$ such that $L_{fu} g = 0$, i.e. $\nabla \vee (f u^\flat) = 0$. In view of (131a) this is equivalent to

$$2\theta + (d \ln f + c^{-2} a^\flat) \vee u^\flat = 0 \,, \tag{140}$$

which, in turn, is equivalent to $\theta = 0$ and $a^\flat = -c^2 \, d \ln f$. This is true since θ is horizontal, $\mathit{\Pi}_h \theta = \theta$, whereas the first term in (140) vanishes upon applying $\mathit{\Pi}_h$. The result now follows from reading this equivalence both ways: 1) The Killing condition for $K := fu$ implies rigidity for u and exactness of $a^\flat$. 2) Rigidity of u and $a^\flat = -d\mathit{\Phi}$ imply that $K := fu$ is Killing, where $f := \exp(\mathit{\Phi}/c^2)$. □

We now return to the condition (139) and express $L_u \omega$ in terms of $du^\flat$. For this we recall that $L_u u^\flat = a^\flat$ (cf. footnote 23) and that Lie derivatives on forms commute with exterior derivatives.[29] Hence we have

$$2 \, L_u \omega = L_u (\mathit{\Pi}_h du^\flat) = \mathit{\Pi}_h da^\flat = da^\flat - c^{-2} u^\flat \wedge L_u a^\flat \,. \tag{141}$$

Here we used the fact that the additional terms that result from the Lie derivative of the projection tensor $\mathit{\Pi}_h$ vanish, as a short calculation shows, and also that on forms the projection tensor $\mathit{\Pi}_h$ can be written as $\mathit{\Pi}_h = \mathrm{id} - c^{-2} u^\flat \wedge i_u$, where i_u denotes the map of insertion of u in the first slot.

Now we prove

Lemma 5. *Let u generate a rigid motion in flat space such that $\omega \neq 0$, then*

$$L_u a^\flat = 0 \,. \tag{142}$$

Proof. Equation (139) says that ω is projectable (it is horizontal by definition). Hence $\hat{\nabla}\omega$ is projectable, which implies

$$L_u \hat{\nabla} \omega = 0 \,. \tag{143}$$

Using (130) with $\theta = 0$ one has

$$\hat{\nabla}\omega = \mathit{\Pi}_h \nabla \omega = \mathit{\Pi}_h \nabla \nabla u^\flat - c^{-2} \mathit{\Pi}_h (\nabla u^\flat \otimes a^\flat) \,. \tag{144}$$

Antisymmetrization in the first two tensor slots makes the first term on the right vanish due to the flatness on ∇. The antisymmetrized right hand side is hence equal to $-c^{-2} \omega \otimes a^\flat$. Taking the Lie derivative of both sides makes the left hand side vanish due to (143), so that

$$L_u (\omega \otimes a^\flat) = \omega \otimes L_u a^\flat = 0 \tag{145}$$

where we also used (139). So we see that $L_u a^\flat = 0$ if $\omega \neq 0$.[30] □

[29] This is most easily seen by recalling that on forms the Lie derivative can be written as $L_u = d \circ i_u + i_u \circ d$, where i_u is the map of inserting u in the first slot.

[30] We will see below that (142) is generally not true if $\omega = 0$; see equation (154).

The last three lemmas now constitute a proof for Theorem 4. Indeed, using (142) in (141) together with (139) shows $da^\flat = 0$, which, according to Lemma 4, implies that the motion is Killing.

Next we turn to the second part of the theorem of Noether and Herglotz, which is somewhat easier to prove:

Theorem 5 (Noether & Herglotz, part 2). *All irrotational rigid motions in Minkowski space are given by the following construction: take a twice continuously differentiable curve $\tau \mapsto z(\tau)$ in Minkowski space, where w.l.o.g τ is the eigentime, so that $\dot{z}^2 = c^2$. Let $H_\tau := z(\tau) + (\dot{z}(\tau))^\perp$ be the hyperplane through $z(\tau)$ intersecting the curve z perpendicularly. Let Ω be a the tubular neighborhood of z in which no two hyperplanes $H_\tau, H_{\tau'}$ intersect for any pair $z(\tau), z(\tau')$ of points on the curve. In Ω define u as the unique (once differentiable) normalized timelike vector field perpendicular to all $H_\tau \cap \Omega$. The flow of u is the sought-for rigid motion.*

Proof. We first show that the flow so defined is indeed rigid, even though this is more or less obvious from its very definition, since we just defined it by "rigidly" moving a hyperplane through spacetime. In any case, analytically we have,

$$H_\tau = \{x \in \mathbb{M}^n \mid f(\tau, x) := \dot{z}(\tau) \cdot \big(x - z(\tau)\big) = 0\}\,. \tag{146}$$

In Ω any x lies on exactly one such hyperplane, H_τ, which means that there is a function $\sigma : \Omega \to \mathbb{R}$ so that $\tau = \sigma(x)$ and hence $F(x) := f(\sigma(x), x) \equiv 0$. This implies $dF = 0$. Using the expression for f from (146) this is equivalent to

$$d\sigma = \dot{z}^\flat \circ \sigma / [c^2 - (\ddot{z} \circ \sigma) \cdot (\mathrm{id} - z \circ \sigma)]\,, \tag{147}$$

where "id" denotes the "identity vector-field", $x \mapsto x^\mu \partial_\mu$, in Minkowski space. Note that in Ω we certainly have $\partial_\tau f(\tau, x) \neq 0$ and hence $\ddot{z} \cdot (x - z) \neq c^2$. In Ω we now define the normalized timelike vector field[31]

$$u := \dot{z} \circ \sigma\,. \tag{148}$$

Using (147), its derivative is given by

$$\nabla u^\flat = d\sigma \otimes (\dot{z}^\flat \circ \sigma) = \big[(\dot{z}^\flat \circ \sigma) \otimes (\ddot{z}^\flat \circ \sigma)\big] / (N^2 c^2)\,, \tag{149}$$

where

$$N := 1 - (\ddot{z} \circ \sigma) \cdot (\mathrm{id} - z \circ \sigma)/c^2\,. \tag{150}$$

This immediately shows that $\Pi_h \nabla u^\flat = 0$ (since $\Pi_h \dot{z}^\flat = 0$) and therefore that $\theta = \omega = 0$. Hence u, as defined in (148), generates an irrotational rigid motion.

For the converse we need to prove that any irrotational rigid motion is obtained by such a construction. So suppose u is a normalized timelike vector field such that $\theta = \omega = 0$. Vanishing ω means $\Pi_h(\nabla \wedge u^\flat) = \Pi_h(du^\flat) = 0$. This is

[31] Note that, by definition of σ, $(\dot{z} \circ \sigma) \cdot (\mathrm{id} - z \circ \sigma) \equiv 0$.

equivalent to $u^\flat \wedge du^\flat = 0$, which according to the Frobenius theorem in differential geometry is equivalent to the integrability of the distribution[32] $u^\flat = 0$, i.e. the hypersurface orthogonality of u. We wish to show that the hypersurfaces orthogonal to u are hyperplanes. To this end consider a spacelike curve $z(s)$, where s is the proper length, running within one hypersurface perpendicular to u. The component of its second s-derivative parallel to the hypersurface is given by (to save notation we now simply write u and $u^\flat$ instead of $u \circ z$ and $u^\flat \circ z$)

$$\Pi_h \ddot{z} = \ddot{z} - c^{-2} u\, u^\flat(\ddot{z}) = \ddot{z} + c^{-2} u\, \theta(\dot{z}, \dot{z}) = \ddot{z}\,, \tag{151}$$

where we made a partial differentiation in the second step and then used $\theta = 0$. Geodesics in the hypersurface are curves whose second derivative with respect to proper length have vanishing components parallel to the hypersurface. Now, (151) implies that geodesics in the hypersurface are geodesics in Minkowski space (the hypersurface is "totally geodesic"), i.e. given by straight lines. Hence the hypersurfaces are hyperplanes. □

Theorem 5 precisely corresponds to the Newtonian counterpart: The irrotational motion of a rigid body is determined by the worldline of any of its points, and any timelike worldline determines such a motion. We can rigidly put an extended body into any state of translational motion, as long as the size of the body is limited by c^2/α, where α is the modulus of its acceleration. This also shows that (142) is generally not valid for irrotational rigid motions. In fact, the acceleration one-form field for (148) is

$$a^\flat = (\ddot{z}^\flat \circ \sigma)/N \tag{152}$$

from which one easily computes

$$da^\flat = (\dot{z}^\flat \circ \sigma) \wedge \left\{ (\Pi_h \dddot{z}^\flat \circ \sigma) + (\ddot{z}^\flat \circ \sigma) \frac{(\Pi_h \ddot{\ddot{z}} \circ \sigma) \cdot (\mathrm{id} - z \circ \sigma)}{Nc^2} \right\} N^{-2} c^{-2}\,. \tag{153}$$

From this one sees, for example, that for *constant acceleration*, defined by $\Pi_h \dddot{z} = 0$ (constant acceleration in time as measured in the instantaneous rest frame), we have $da^\flat = 0$ and hence a Killing motion. Clearly, this is just the motion (124) for the boost Killing field (123). The Lie derivative of $a^\flat$ is now easily obtained:

$$L_u a^\flat = i_u da^\flat = (\Pi_h \ddot{z}^\flat \circ \sigma) N^{-2}\,, \tag{154}$$

showing explicitly that it is not zero except for motions of constant acceleration, which were just seen to be Killing motions.

[32] "Distribution" is here used in the differential-geometric sense, where for a manifold M it denotes an assignment of a linear subspace V_p in the tangent space T_pM to each point p of M. The distribution $u^\flat = 0$ is defined by $V_p = \{v \in T_pM \mid u_p^\flat(v) = u_p \cdot v = 0\}$. A distribution is called (locally) integrable if (in the neighborhood of each point) there is a submanifold M' of M whose tangent space at any $p \in M'$ is just V_p.

In contrast to the irrotational case just discussed, we have seen that we cannot put a body rigidly into rotational motion. In the old days this was sometimes expressed by saying that the rigid body in SR has only three instead of six degrees of freedom. This was clearly thought to be paradoxical as long as one assumed that the notion of a perfectly rigid body should also make sense in the framework of SR. However, this hope was soon realized to be physically untenable [36].

5.6 Geometry of Space and Time in Rotating Reference Frames

We have seen above that there is a generalization of Einstein simultaneity for the case of rigid linear accelerations. The hypersurfaces of simultaneity were given by the hyperplanes of constant Killing time λ, which are different from the (curved) hypersurfaces of constant proper time τ. This worked because the Killing field was (locally) hypersurface orthogonal.

Note that in terms of the co-rotating coordinates (ct, z, ρ, ψ) (recall that $\psi = \varphi - \kappa t$) the Killing field (133) is just $K = \partial_t$. It is convenient to rewrite the spacetime metric $g = c^{-2} u^\flat \otimes u^\flat - h$ in the following form

$$g = c^2 \, \exp(2\Phi/c^2) \; A \otimes A - h \,, \tag{155}$$

where h is given by (135b) and, using (135a), we have the following expressions for Φ and A:

$$\Phi \; := \; c^2 \, \ln \left\{ \sqrt{K^2/c^2} \right\} \; = \; \tfrac{c^2}{2} \ln \left\{ 1 - (\kappa\rho/c)^2 \right\} \,, \tag{156a}$$

$$A \; := \qquad K^\flat / K^2 \qquad = \; dt - \frac{\kappa\rho^2/c^2}{1 - (\kappa\rho/c)^2} \, d\psi \,. \tag{156b}$$

The physical interpretation of Φ appears from calculating the acceleration a that an observer experiences who moves along the Killing orbit:

$$a^\flat := \nabla_u u^\flat = -d\Phi \,. \tag{157}$$

Hence Φ is the Newtonian potential that is accelerating the Killing observer.

The rotational Killing field is clearly not hypersurface orthogonal. The obstruction is just given by the vorticity ω. A simple calculation gives

$$F := dA = 2c^{-2} \exp(-\Phi/c^2) \, \omega \,. \tag{158}$$

Hence the obstruction for hypersurface orthogonality is likewise faithfully measured by A. Moreover, as we shall see below, the 1-form A has an interesting physical and geometric interpretation, which is the actual reason why we introduced it here.

Inside the region Ω (defined in (134)) K is a complete and nowhere vanishing timelike vector field. This means that the flow $f : \mathbb{R} \times \Omega \to \Omega$ of K defines a free action of the additive group $\mathbb{R}$ on Ω that makes Ω the total space of a principle bundle with fiber $\mathbb{R}$ and base $\hat{\Omega} = \Omega/\!\sim$. Here $\sim$ is again the equivalence

relation which declares two points in Ω to be equivalent iff they lie on the same K orbit. Hence $\hat{\Omega}$, which is obviously diffeomorphic to the solid cylinder $\{(z, \rho, \varphi) \mid \rho < c/\kappa\}$, is the space of K orbits. Since Ω is endowed with the metric g and since K acts by isometries, the distribution of hyperplanes (117) orthogonal to the Killing orbits define a connection on the principal bundle whose corresponding 1-form is just A.[33] Accordingly, the bundle curvature is given by $F = dA$. Note that F can be considered as 2-form on $\hat{\Omega}$ since $i_K F = 0$ and $L_K F = i_K dF = 0$.[34]

Now, parallel transport defined by the connection A has a direct physical interpretation: it is just transportation of time according to Einstein synchronization. Since $F \neq 0$ this transportation is not path independent. In particular this implies that synchronization along fixed paths is not a transitive operation anymore. Given two points in $\hat{\Omega}$ connected by a spatial path $\hat{\gamma}$ in $\hat{\Omega}$, parallel transportation along $\hat{\gamma}$ requires that we lift $\hat{\gamma}$ to a path γ in Ω whose tangent vectors are annihilated by A, that is, which runs orthogonally to the orbits of K. But this is just what we mean by saying that the points on the curve γ are locally Einstein synchronized, in the sense that any two infinitesimally nearby points on γ are Einstein synchronized. Hence the integral of A along γ vanishes. Using (156b) this is equivalent to

$$\Delta t := \int_\gamma dt = \frac{\kappa}{c^2} \int_{\hat{\gamma}} \frac{\rho^2}{1 - (\kappa\rho/c)^2}\, d\psi\,, \tag{159}$$

where we interpreted ρ and ψ as coordinates on $\hat{\Omega}$ so that the right hand side could be written as integral along the curve $\hat{\gamma}$ in $\hat{\Omega}$. This means that if we Einstein synchronize clocks along $\hat{\gamma}$ in space, the clock at the final point of $\hat{\gamma}$ shows a lapse Δt of coordinate-time as compared to the clock at the initial point of $\hat{\gamma}$. A striking consequence of the non-transitivity of Einstein synchronization is the non-zero lapse of coordinate time that one obtains for spatially closed curves. These lapses are just the holonomies of the connection A. If, for simplicity, we choose $\hat{\gamma}$ to be a closed planar loop of constant ρ and z, (159) immediately leads to

$$\Delta t = \frac{2\pi}{\kappa} \frac{(\kappa\rho/c)^2}{1 - (\kappa\rho/c)^2} \approx (2\kappa/c^2)\mathcal{S}\,, \tag{160}$$

where $\mathcal{S}$ is explained below. The lapse in proper time, $\Delta\tau$, is obtained by multiplying this result with $\exp(\Phi/c^2)$, which merely amounts to replacing the denominator $1 - (\kappa\rho/c)^2$ in (160) with its square root. This time lapse is directly related to the *Sagnac effect*. In fact, the observed phase shift in the Sagnac

[33] The connection 1-form A associated to the distribution of "horizontal" subspaces has to fulfill two conditions: 1) vectors tangential to the horizontal subspaces are annihilated by A and 2) $A(K) = 1$, where K is a "fundamental vector field" which generates the action of the structure group $\mathbb{R}$. Both conditions are satisfied in our case. See e.g. [9] for a lucid discussion of these notions.

[34] More precisely, there is a unique 2-form $\hat{F}$ on $\hat{\Omega}$ such that $\pi^* \hat{F} = F$, where $\pi : \Omega \to \hat{\Omega}$ is the bundle projection.

effect is obtained by multiplying the expression for the time lapse with twice[35] the light's frequency ν.

In (160) $\mathcal{S}$ denotes the area of the 2-disk spanned by the planar loop. Note that this area is only approximately given by $\pi\rho^2$ since the geometry in $\hat{\Omega}$, determined with co-rotating rods and clocks, is given by the metric h; see (135b). The precise expressions for the circumference, $\mathcal{C}$ and area, $\mathcal{S}$, of the planar loop of constant ρ follow from (135b):

$$\mathcal{C} = \int_0^{2\pi} \frac{d\psi\,\rho}{\sqrt{1-(\kappa\rho/c)^2}} = \frac{2\pi\rho}{\sqrt{1-(\kappa\rho/c)^2}} > 2\pi\rho\,, \quad (161a)$$

$$\mathcal{S} = \int_0^{2\pi}\int_0^{\rho} \frac{d\psi\,d\rho'\,\rho'}{\sqrt{1-(\kappa\rho'/c)^2}} = \frac{2\pi c^2}{\kappa^2}\left\{1-\sqrt{1-(\kappa\rho/c)^2}\right\} > \pi\rho^2\,. \quad (161b)$$

The circumference grows faster than $\propto \rho$ and the area faster than $\propto \rho^2$. Note that according to (135b) ρ is the geodesic radial distance. Hence the two-dimensional hypersurfaces of constant z in $\hat{\Omega}$ are negatively curved. In fact, the Gaussian curvature, $\mathcal{K}$, of these hypersurfaces turns out to be

$$\mathcal{K} = \frac{-3\,(\kappa/c)^2}{\left\{1-(\kappa\rho/c)^2\right\}^2}\,, \quad (162)$$

which is strictly negative, approximately constant for $\rho \ll c/\kappa$, and unbounded as ρ approaches the critical radius c/κ. In contrast, according to (135b), the metrics induced by h on the hypersurfaces of constant ψ are flat.

The bundle curvature of F for the connection A and the Riemannian curvature for $(\hat{\Omega}, h)$ are indeed intimately linked through identities which arise by calculating the Riemannian curvature of (Ω, g), where g is parameterized as in (155), and noting that g is flat (Minkowski metric). One such identity is the so called *Kaluza-Klein identity*, which expresses the scalar curvature (Ricci scalar) of g in terms of the scalar curvature R_h of h, Φ, and $\|F\|_h^2 = h^{ik}h^{jl}F_{ij}F_{kl}$. Since the scalar curvature of g is zero, one obtains:

$$R_h = 2\exp(-\Phi/c^2)\Delta_h \exp(\Phi/c^2) - \tfrac{1}{4}\,c^2\exp(2\Phi/c^2)\,\|F\|_h^2\,, \quad (163)$$

where Δ_h is the Laplace operator on $(\hat{\Omega}, h)$.

It is an interesting historical fact that it was Kaluza who pointed out that "space" in rotating reference frames cannot be identified with a submanifold perpendicular to the Killing orbits (because such a submanifold does not exist) but rather has to be constructed as the quotient manifold $\hat{\Omega}$ which carries the curved metric h [34]. He also discussed the non-integrability of Einstein synchronization. This he did in 1910, ten years before he applied the very same mathematical ideas to the five-dimensional setting known as Kaluza-Klein theories.

[35] The factor 2 results simply from the fact that the Sagnac effect measures the *sum* of the moduli of time lapses for a closed curve traversed in both directions.

A Appendices

For the interested reader this appendix collects some mathematical background material which are relevant to the discussion in the main text.

A.1 Sets and Group Actions

Given a set S, recall that an *equivalence relation* is a subset $R \subset S \times S$ such that for all $p, q, r \in S$ the following conditions hold: 1) $(p, p) \in R$ (called "reflexivity"), 2) if $(p, q) \in R$ then $(q, p) \in R$ (called "symmetry"), and 3) if $(p, q) \in R$ and $(q, r) \in R$ then $(p, r) \in R$ (called "transitivity"). Once R is given, one often conveniently writes $p \sim q$ instead of $(p, q) \in R$. Given $p \in S$, its *equivalence class*, $[p] \subseteq S$, is given by all points R-related to p, i.e. $[p] := \{q \in S \mid (p, q) \in R\}$. One easily shows that equivalence classes are either identical or disjoint. Hence they form a *partition* of S, that is, a covering by mutually disjoint subsets. Conversely, given a partition of a set S, it defines an equivalence relation by declaring two points as related iff they are members of the same cover set. Hence there is a bijective correspondence between partitions of and equivalence relations on a set S. The set of equivalence classes is denoted by S/R or $S/\sim$. There is a natural surjection $S \to S/R$, $p \mapsto [p]$.

If in the definition of equivalence relation we exchange symmetry for antisymmetry, i.e. $(p, q) \in R$ and $(q, p) \in R$ implies $p = q$, the relation is called a *partial order*, usually written as $p \geq q$ for $(p, q) \in R$. If, instead, reflexivity is dropped and symmetry is replaced by asymmetry, i.e. $(p, q) \in R$ implies $(q, p) \notin R$, one obtains a relation called a *strict partial order*, usually denoted by $p > q$ for $(p, q) \in R$.

An *left action* of a group G on a set S is a map $\phi : G \times S \to S$, such that $\phi(e, s) = s$ (e = group identity) and $\phi(gh, s) = \phi(g, \phi(h, s))$. If instead of the latter equation we have $\phi(gh, s) = \phi(h, \phi(g, s))$ one speaks of a *right action*. For left actions one sometimes conveniently writes $\phi(g, s) =: g \cdot s$, for right actions $\phi(g, s) =: s \cdot g$. An action is called *transitive* if for every pair $(s, s') \in S \times S$ there is a $g \in G$ such that $\phi(g, s) = s'$, and *simply transitive* if, in addition, (s, s') determine g uniquely, that is, $\phi(g, s) = \phi(g', s)$ for some s implies $g = g'$. The action is called *effective* if $\phi(g, s) = s$ for all s implies $g = e$ ('every $g \neq e$ moves something') and *free* if $\phi(g, s) = s$ for some s implies $g = e$ ('no $g \neq e$ has a fixed point'). It is obvious that simple transitivity implies freeness and that, conversely, freeness and transitivity implies simple transitivity. Moreover, for Abelian groups, effectivity and transitivity suffice to imply simple transitivity. Indeed, suppose $g \cdot s = g' \cdot s$ holds for some $s \in S$, then we also have $k \cdot (g \cdot s) = k \cdot (g' \cdot s)$ for all $k \in G$ and hence $g \cdot (k \cdot s) = g' \cdot (k \cdot s)$ by commutativity. This implies that $g \cdot s = g' \cdot s$ holds, in fact, for all s.

For any $s \in S$ we can consider the *stabilizer subgroup*

$$\mathrm{Stab}(s) := \{g \in G \mid \phi(g, s) = s\} \subseteq G\,. \tag{164}$$

If ϕ is transitive, any two stabilizer subgroups are conjugate: $\mathrm{Stab}(g \cdot s) = g\mathrm{Stab}(s)g^{-1}$. By definition, if ϕ is free all stabilizer subgroups are trivial (consist

of the identity element only). In general, the intersection $G' := \bigcap_{s\in S} \mathrm{Stab}(s) \subseteq G$ is the normal subgroup of elements acting trivially on S. If ϕ is an action of G on S, then there is an effective action $\hat{\phi}$ of $\hat{G} := G/G'$ on S, defined by $\hat{\phi}([g], s) := \phi(g, s)$, where $[g]$ denotes the G'-coset of G' in G.

The *orbit* of s in S under the action ϕ of G is the subset

$$\mathrm{Orb}(s) := \{\phi(g, s) \mid g \in G\} \subseteq S\,. \tag{165}$$

It is easy to see that group orbits are either disjoint or identical. Hence they define a partition of S, that is, an equivalence relation.

A relation R on S is said to be invariant under the self map $f : S \to S$ if $(p, q) \in R \Leftrightarrow (f(p), f(q)) \in R$. It is said to be invariant under the action ϕ of G on S if $(p, q) \in R \Leftrightarrow (\phi(g, p), \phi(g, q)) \in R$ for all $g \in G$. If R is such a G-invariant equivalence relation, there is an action ϕ' of G on the set S/R of equivalence classes, defined by $\phi'(g, [p]) := [\phi(g, p)]$. A general theorem states that invariant equivalence relations exist for transitive group actions, iff the stabilizer subgroups (which in the transitive case are all conjugate) are maximal (e.g. Theorem 1.12 in [32]).

A.2 Structures on Vector and Affine Spaces

A.3 Non Degenerate Bilinear Forms

Consider a vector space V of dimension n over $\mathbb{F}$ (here denoting $\mathbb{R}$ or $\mathbb{C}$). Let it be endowed with a non-degenerate bilinear form $\omega : V \times V \to \mathbb{F}$. No assumptions regarding symmetries of ω are made at this point. The dual space of V is denoted by V^* whose elements we will denote by Greek letters. The set of linear maps $V \to V$ is denoted by $\mathrm{End}(V)$, called the endomorphisms of V, which forms an associative algebra over $\mathbb{F}$ (algebra multiplication being composition of maps). The set of invertible elements in $\mathrm{End}(V)$ (i.e. isomorphisms of V) will be denoted by $\mathsf{GL}(V)$; it forms a group under composition. Generally, composition of maps will be denoted by $\circ$.

The form ω defines an isomorphism

$$\omega^\downarrow : V \to V^*\,, \qquad \omega^\downarrow(v) := \omega(v, \cdot)\,, \tag{166}$$

with inverse map being denoted by

$$\omega^\uparrow : V^* \to V\,, \qquad \omega^\uparrow := (\omega^\downarrow)^{-1}\,, \tag{167}$$

so that

$$\omega^\uparrow \circ \omega^\downarrow = \mathrm{id}_V \qquad \text{and} \qquad \omega^\downarrow \circ \omega^\uparrow = \mathrm{id}_{V^*}\,. \tag{168}$$

Recall that "transposition" is a map $\mathrm{End}(V) \to \mathrm{End}(V^*)$, denoted by $A \mapsto A^\top$ and defined through $A^\top(\alpha) := \alpha \circ A$. This map is an anti-isomorphism of algebras ('anti', since it obeys $(A \circ B)^\top = B^\top \circ A^\top$). Different from this canonically defined notion of transposition is the 'ω-transposition', which is an

isomorphism $\mathrm{End}(V) \to \mathrm{End}(V)$, which we denote by A^t (the dependence on ω being implicitly understood) and which is defined through

$$\omega(A^t u, v) = \omega(u, Av) \quad \forall u, v \in V\,. \tag{169}$$

Note that the ω-transposed is in $\mathrm{End}(V)$ whereas the canonical transposed is in $\mathrm{End}(V^*)$. The relations between the two are

$$A^t = \omega^\uparrow \circ A^\top \circ \omega^\downarrow \quad \text{and} \quad A^\top = \omega^\downarrow \circ A^t \circ \omega^\uparrow\,. \tag{170}$$

A.4 Generalized Orthogonal Transformations

A generalized orthogonal transformation of (V, ω) is any bijective map $\phi : V \to V$ such that $\omega\big(\phi(u), \phi(v)\big) = \omega(u, v)$ for all $u, v \in V$. In this subsection we shall restrict to symmetric ω. Note that any symmetric bilinear form ω is uniquely determined by its quadratic form, i.e. the function $\hat{\omega} : V \to \mathbb{F}$, $v \mapsto \hat{\omega}(v) := \omega(v, v)$, for we have $\omega(u, v) = \frac{1}{2}\big(\hat{\omega}(u+v) - \hat{\omega}(u) - \hat{\omega}(v)\big)$. It is sometimes useful to consider generalizations of distance measures by setting $d(u, v) := \sqrt{|\hat{\omega}(u - v)|}$. This is e.g. done in SR, where one speaks of timelike and spacelike distances in that sense. Now suppose φ is an isometry with respect to d, i.e. $d(\varphi(u), \varphi(v)) = d(u, v)$ for all u, v. Consider ϕ defined by $\phi(u) := \varphi(u) - \varphi(0)$. Clearly φ is an isometry of d if ϕ is an orthogonal transformation with respect to ω. Now, orthogonal transformations are necessarily linear:

Proposition 9. *Let ω be a non-degenerate symmetric bilinear form on V and let $\phi : V \to V$ be an orthogonal transformation with respect to ω. Then ϕ is linear.*

Proof. Consider $I := \omega\big(a\phi(u) + b\phi(v) - \phi(au + bv), w\big)$; surjectivity[36] allows to write $w = \phi(z)$, so that $I = a\omega(u, z) + b\omega(v, z) - \omega(au + bv, z) = 0$ for all $z \in V$. Hence the aforementioned expression for I is zero for all $w \in V$, which by non-degeneracy of ω implies the linearity of ϕ. □

Particularly simple orthogonal transformations are given by reflections on non-degenerate hyperplanes. To explain this, let $v \in V$ and $v^\perp := \{w \in V \mid \omega(v, w) = 0\} \subset V$. $v^\perp$ is a linear subspace of co-dimension one, that is, a hyperplane. That it be non-degenerate means that $\omega|_{v^\perp}$ is non-degenerated, which is easily seen to be the case iff $\omega(v, v) \neq 0$. The reflection at the non-degenerate hyperplane $v^\perp$ is the map

$$\rho_v(x) := x - 2\,v\,\frac{v \cdot x}{v^2}\,. \tag{171}$$

where for convenience we wrote $u \cdot v := \omega(u, v)$ and $v^2 := v \cdot v$. ρ_v is easily seen to be an involutive (i.e. $\rho_v \circ \rho_v = \mathrm{id}_V$) orthogonal transformation. If ϕ is any other orthogonal transformation, the following equivariance property holds: $\phi \circ \rho_v \circ \phi^{-1} = \rho_{\phi(v)}$. An important result is now given by

[36] Note that we only use surjectivity here, so that the hypotheses for this result may be slightly reduced.

Theorem 6 (Cartan, Dieudonné). *Let the dimension of V be n. Any orthogonal transformation of (V, ω) is the composition of at most n reflections.*

Proof. Comprehensive proofs may be found in [32] or [7]. Here we offer a proof of the weaker result, that any orthogonal transformation is the composition of at most $2n-1$ reflections. So let ϕ be orthogonal and $v \in V$ so that $v^2 \neq 0$ (which certainly exists). Let $w = \phi(v)$, then $(v+w)^2 + (v-w)^2 = 4v^2 \neq 0$ so that $w+v$ and $w-v$ cannot simultaneously have zero squares. So let $(v \mp w)^2 \neq 0$ (understood as alternatives), then $\rho_{v \mp w}(v) = \pm w$ and $\rho_{v \mp w}(w) = \pm v$. Hence v is eigenvector with eigenvalue 1 of the orthogonal transformation given by

$$\phi' = \begin{cases} \rho_{v-w} \circ \phi & \text{if } (v-w)^2 \neq 0\,, \\ \rho_v \circ \rho_{v+w} \circ \phi & \text{if } (v-w)^2 = 0\,. \end{cases} \tag{172}$$

Consider now the orthogonal transformation $\phi'\big|_{v^\perp}$ on $v^\perp$ with induced bilinear form $\omega\big|_{v^\perp}$, which is non-degenerated due to $v^2 \neq 0$. We now conclude by induction: At each dimension we need at most two reflections to reduce the problem by one dimension. After $n-1$ steps we have reduced the problem to one dimension, where we need at most one more reflection. Hence we need at most $2(n-1)+1 = 2n-1$ reflections which upon composition with ϕ produce the identity. Here we use that any orthogonal transformation in $v^\perp$ can be canonically extended to $\mathrm{span}\{v\} \oplus v^\perp$ by just letting it act trivially on $\mathrm{span}\{v\}$. □

There are several useful applications of this result, most notably in the construction of the Spin groups. Other applications in SR are discussed in [48].

A.5 Index Raising and Lowering

Let $\{e_a\}_{a=1,\cdots,n}$ be a basis of V and $\{\eta^a\}_{a=1,\cdots,n}$ its *(canonical) dual basis* of V^*, which is defined by $\eta^a(e_b) = \delta^a_b$. Using $\omega^\downarrow$ and $\omega^\uparrow$ one can define the ω-duals of $\{e_a\}$ and $\{\eta^a\}$ respectively, given by

$$\eta_a := \omega^\downarrow(e_a) \in V^*\,, \tag{173a}$$

$$e^a := \omega^\uparrow(\eta^a) \in V\ \ , \tag{173b}$$

so that, writing $\omega_{ab} := (e_a, e_b)$ and ω^{ab} for the components of the inverse-transposed matrix (i.e. $\omega_{ac}\omega^{bc} = \omega_{ca}\omega^{cb} = \delta^b_a$),

$$\eta_a := \omega^\downarrow(e_a) = \omega_{ab}\eta^b\,, \tag{174a}$$

$$e^a := \omega^\uparrow(\eta^a) = \omega^{ba}e_b\,. \tag{174b}$$

Using components with respect to the canonical dual bases, so that $v = v^a e_a \in V$ with $\omega^\downarrow(v) =: v_a\eta^a$ and $\alpha = \alpha_a\eta^a \in V^*$ with $\omega^\uparrow(\alpha) =: \alpha^a e_a$, one obtains the equivalent to (174) in coordinates:

$$v_a = v^b\omega_{ba}\,, \tag{175a}$$

$$\alpha^a = \omega^{ab}\alpha_b\,. \tag{175b}$$

It should be clear from (174) and (175) why the maps $\omega^\downarrow$ and $\omega^\uparrow$ are called "index lowering" and "index raising". Often, if there is no ambiguity as to what structure ω is used, the following notation is employed: Let $v \in V$ and $\alpha \in V^*$, then $\omega^\downarrow(v) =: v^\flat \in V^*$ and $\omega^\uparrow(\alpha) =: \alpha^\sharp \in V$.

Finally we remark on the choice of conventions. Comparing e.g. (175a) with (175b) one notices that one sums over the first index on ω for lowering and over the second index for raising indices (on the bases (174) it is just the other way round). This is a consequence of the following requirements: 1) raising and lowering of indices are mutually inverse operations, and 2) the matrix, $\{\omega^{ab}\}$, used for raising indices is the *transposed* inverse of $\{\omega_{ab}\}$. The rationale for the second condition is the requirement that lowering both indices on $\{\omega^{ab}\}$ using $\{\omega_{ab}\}$ should reproduce $\{\omega_{ab}\}$ and raising both indices on $\{\omega_{ab}\}$ using $\{\omega^{ab}\}$ should reproduce $\{\omega^{ab}\}$. This enforces 2).

Note again that so far no assumptions were made concerning the symmetries of ω. In the general case there are, in fact, two raising-lowering operations: One as given above, the other by replacing (166) with $\tilde{\omega}^\downarrow(v) := \omega(\cdot, v)$, i.e. v now being in the second rather than the first slot. For this second operation we have all formulae as above with $\{\omega_{ab}\}$ and $\{\omega^{ab}\}$ being replaced by the transposed matrices. In physical applications ω is either symmetric – like in case of the Minkowski metric – or antisymmetric – like for the 2-spinor metric (symplectic form). In those cases there is – up to sign in the second case – a unique pair of lowering and raising operations.

A.6 Linear Frames

A basis $f = \{e_a\}_{a=1,\cdots,n}$ of V can be viewed as a linear isomorphism (also denoted by f), $f : \mathbb{F}^n \to V$, given by $f(v^1, \cdots, v^n) = v^a e_a$. With this interpretation we call the basis f a *linear frame*. Any frame f induces an isomorphism of algebras $\mathrm{End}(\mathbb{F}^n) \to \mathrm{End}(V)$, given by $A \mapsto A^f := f \circ A \circ f^{-1}$. If $A = \{A^b_a\}$, then $A^f(e_a) = A^b_a e_b$. The standard (linear) action ϕ of $\mathsf{GL}(\mathbb{F}^n)$ on $\mathbb{F}^n$, $\phi(A, x) := Ax$, thereby translates in an f-dependent way to an action ϕ^f of $\mathsf{GL}(\mathbb{F}^n)$ on V, defined by $\phi^f(A, v) := f \circ \phi(A, f^{-1}(v))$; that is, $(A, v) \mapsto A^f v = f(Ax)$, where $f(x) = v$.

Let $\mathcal{F}_V$ denote the set of frames for V. The *general linear group* $\mathsf{GL}(\mathbb{F}^n)$ acts transitively and freely on $\mathcal{F}_V$ from the right:

$$\mathsf{GL}(V) \times \mathcal{F}_V \to \mathcal{F}_V\,, \qquad (A, f) \mapsto f \cdot A := f \circ A\,. \tag{176}$$

Proper subgroups of $\mathsf{GL}(\mathbb{F}^n)$ continue to act freely on $\mathcal{F}_V$.

A.7 Affine Spaces

An affine space over the vector space V is a set $\mathrm{Aff}(V)$ together with an effective and transitive action ϕ of V, considered as Abelian group (group multiplication being vector addition). Since the group is Abelian, this suffices to imply that the action is free and simply transitive. One writes $\phi(m, v) =: m + v$, which

defines what is meant by "+" between an element of $\mathrm{Aff}(V)$ and an element of V. Any ordered pair of points $(p, q) \in \mathrm{Aff}(V) \times \mathrm{Aff}(V)$ uniquely defines a vector v, namely that for which $p = q + v$. One writes $p - q = v$, defining what is meant by "−" between two elements of $\mathrm{Aff}(V)$. Considered as Abelian groups, any linear subspace $W \subset V$ defines a subgroup. The orbit of that subgroup in $\mathrm{Aff}(V)$ through $m \in \mathrm{Aff}(V)$ is an affine subspace, denoted by W_m, i.e.

$$W_m = m + W := \{m + w \mid w \in W\}\,, \tag{177}$$

which is an affine space over W in its own right of dimension $\dim(W)$. One-dimensional affine subspaces are called *(straight) lines*, two-dimensional ones *planes*, and those of co-dimension one are called *hyperplanes*.

A.8 Affine Frames

A *basis* for $\mathrm{Aff}(V)$ is a tuple $F := (m, f)$, where m is a point in $\mathrm{Aff}(V)$ and f a basis of V. F can be considered as a map $\mathbb{F}^n \to \mathrm{Aff}(V)$, given by $F(x) := f(x) + m$ (here f is interpreted as linear frame). With this interpretation F is called an *affine frame*. We denote the set of affine frames by $\mathcal{F}_{\mathrm{Aff}(V)}$.

The *general affine group* of $\mathbb{F}^n$ is given in the familiar fashion by the semi-direct product $\mathbb{F}^n \rtimes \mathsf{GL}(\mathbb{F}^n)$, which acts on $\mathbb{F}^n$ in the standard way: $\phi : ((a, A), x) \mapsto \phi((a, A), x) := A(x) + a$. Its multiplication law is given by:

$$(a_1, A_1)(a_2, A_2) = (a_1 + A_1 a_2\,,\; A_1 A_2)\,. \tag{178}$$

Depending on the choice of a frame $F \in \mathcal{F}_{\mathrm{Aff}(V)}$ the action ϕ of $\mathbb{F}^n \rtimes \mathsf{GL}(\mathbb{F}^n)$ on $\mathbb{F}^n$ translates to an action ϕ^F of $\mathbb{F}^n \rtimes \mathsf{GL}(\mathbb{F}^n)$ on $\mathrm{Aff}(V)$ as follows: $\phi^F((a, A), p) := F \circ \phi((a, A), F^{-1}(p))$; in other words, if $F = (m, f)$ and $F(x) = p$, we have $\phi^F : ((a, A), p) \mapsto F(Ax + a) = A^f(p - m) + m + f(a)$.

$\mathbb{F}^n \rtimes \mathsf{GL}(\mathbb{F}^n)$ has an obvious right action on $\mathcal{F}_{\mathrm{Aff}(V)}$, given by $(g, F) \mapsto F \cdot g := F \circ g$. Explicitly, for $g = (a, A)$ and $F = (m, f)$, this reads

$$F \cdot g = (m, f) \cdot (a, A) = (m + f(a), f \circ A)\,. \tag{179}$$

A.9 Lie Algebras for Matrix Groups

General Considerations

We first recall the definition of a Lie algebra:

Definition 8. *A* Lie algebra *over* $\mathbb{F}$ *(here denoting* $\mathbb{R}$ *or* $\mathbb{C}$*) is a vector space* L *over* $\mathbb{F}$ *endowed with a map (called the "Lie bracket")* $L \times L \to L$, $(X, Y) \mapsto [X, Y]$, *which for all* $X, Y, Z \in L$ *obeys:*

$$[X, Y] = -[Y, X] \qquad \textit{(anti-symmetry)}\,, \tag{180a}$$

$$[aX + Y, Z] = a[X, Z] + [Y, Z] \qquad \textit{(linearity)}\,, \tag{180b}$$

$$[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0 \qquad \textit{(Jacobi identity)}\,. \tag{180c}$$

A Lie *subalgebra* $L' \subseteq L$ is a linear subspace which becomes a Lie algebra when the bracket is restricted to L', i.e. if $[L', L'] \subseteq L'$. A Lie subalgebra is called an *ideal* if the stronger condition holds that $[L', L] \subseteq L'$. It is easy to see that if L' is an ideal the quotient L/L' is again a Lie algebra: just define the bracket of two cosets as the coset of the bracket of two arbitrary representatives, which is well defined.

In may cases of interest L is already given as an associative algebra and the Lie bracket is then defined as commutator: $[X, Y] := X \cdot Y - Y \cdot X$. This is e.g. the case if $L \subseteq \mathrm{End}(V)$ since, as already mentioned, the endomorphisms of a vector space V form an associative algebra if the multiplication is taken to be the composition of maps.

Given a matrix group $\mathsf{G} \subseteq \mathsf{GL}(\mathsf{n}, \mathbb{F})$ we consider the set $C^1_*(\mathbb{R}, \mathsf{G})$ of all continuously differentiable curves $A : \mathbb{R} \to \mathsf{G}$ such that $A(0) = \mathbf{1}_n$ (unit $n \times n$-matrix). We define $\dot{A} := \frac{d}{ds} A(s)|_{s=0}$, the "velocity" of the curve $A(s)$ at the group identity. We consider the set of all such velocities:

$$\mathrm{Lie}(\mathsf{G}) := \left\{ \dot{A} \;\middle|\; A \in C^1_*(\mathbb{R}, \mathsf{G}) \right\} \subset \mathrm{End}(\mathbb{R}^n)\,. \tag{181}$$

Proposition 10. $\mathrm{Lie}(\mathsf{G})$ *is a real Lie algebra.*

Proof. First we prove that $\mathrm{Lie}(\mathsf{G})$ is a linear space: Let $X, Y \in \mathrm{Lie}(\mathsf{G})$ and $A, B \in C^1_*(\mathbb{R}, \mathsf{G}))$ such that $X = \dot{A}$ and $Y = \dot{B}$. Define $C \in C^1_*(\mathbb{R}, \mathsf{G})$ by $C(s) := A(s) \cdot B(ks)$, where $k \in \mathbb{R}$, then $\dot{C} = X + kY$, showing that $\mathrm{Lie}(\mathsf{G})$ is a vector space over $\mathbb{R}$. Here and below "$\cdot$" denotes matrix multiplication. Now, since $\mathrm{Lie}(\mathsf{G}) \subseteq \mathrm{End}(\mathbb{F}^n)$, i.e. lies in an associative algebra, we define the Lie bracket on $\mathrm{Lie}(\mathsf{G})$ as commutator, that is $[X, Y] := X \cdot Y - X \cdot X$. This bracket clearly satisfies conditions (180)). But we still have to show that $[X, Y]$ is in $\mathrm{Lie}(\mathsf{G})$ if X, Y are. That is, we have to show that there is a curve $C \in C^1_*(\mathbb{R}, \mathsf{G})$ such that $\dot{C} = [X, Y]$. To do this, let again $A, B \in C^1_*(\mathbb{R}, \mathsf{G})$ be such that $X = \dot{A}$ and $Y = \dot{B}$. Then the sought for C is given by

$$C(s) := \begin{cases} A(\tau(s)) \cdot B(\tau(s)) \cdot A^{-1}(\tau(s)) \cdot B^{-1}(\tau(s)) & \text{for } s \geq 0\,, \\ B(\tau(s)) \cdot A(\tau(s)) \cdot B^{-1}(\tau(s)) \cdot A^{-1}(\tau(s)) & \text{for } s \leq 0\,, \end{cases} \tag{182}$$

where

$$\tau(s) := \begin{cases} \sqrt{s} & \text{for } s \geq 0\,, \\ -\sqrt{-s} & \text{for } s \leq 0\,. \end{cases} \tag{183}$$

This curve is indeed differentiable at $s = 0$ (though $s \mapsto A(\sqrt{s})$ and $s \mapsto B(\sqrt{s})$ are not). Its right derivative ($s \geq 0$) is:

$$\begin{aligned} \dot{C} &= \lim_{s \to 0} \frac{C(s) - \mathbf{1}_n}{s} = \lim_{s \to 0} \left\{ \frac{\big[A(\tau(s)), B(\tau(s))\big] A^{-1}(\tau(s)) B^{-1}(\tau(s))}{s} \right\} \\ &= \lim_{\tau \to 0} \left\{ \left[\frac{A(\tau) - \mathbf{1}_n}{\tau}\,,\, \frac{B(\tau) - \mathbf{1}_n}{\tau} \right] A^{-1}(\tau) B^{-1}(\tau) \right\} = [X, Y]\,. \end{aligned} \tag{184}$$

Its left derivative follows along the same lines, one just exchanges $A \leftrightarrow B$ and replaces s with $-s$, leading again to $[X, Y]$. □

Some Special Lie Algebras

Before we restrict attention to the Lorentz group and its inhomogeneous counterpart (sometimes called the Poincaré group), let us describe in general the situation of which they are special cases.

Consider a vector space V of n dimensions over the field $\mathbb{F}$ ($\mathbb{R}$ or $\mathbb{C}$). As before, $\mathrm{End}(V)$ denotes the associative algebra of linear maps $V \mapsto V$. Let $\mathsf{GL}(V) \subset \mathrm{End}(V)$ denote the set of invertible linear maps, i.e. $\det(f) \neq 0$ (compare (195)) for all $f \in \mathsf{GL}(V)$.

Given a subgroup $\mathsf{G} \subseteq \mathsf{GL}(V)$, there is a corresponding inhomogeneous group, $\mathsf{IG} \subseteq \mathsf{IGL}(V)$, given by the semi-direct product of V (considered as Abelian group under addition) with G, denoted by $V \rtimes \mathsf{G}$. Its multiplication law is as follows:

$$(a_1, A_1)(a_2, A_2) = (a_1 + A_1(a_2)\,,\, A_1 \circ A_2)\,, \qquad a_i \in V \;\; A_i \in \mathsf{G} \tag{185}$$

We endow V with a non-degenerate bilinear form $\omega : V \times V \to \mathbb{F}$, which we restrict to be either symmetric ($\epsilon = 1$) or antisymmetric ($\epsilon = -1$), that is $\omega(v, w) = \epsilon\,\omega(w, v)$ for all $v, w \in V$. We want to consider the group $\mathsf{G} \subset \mathsf{GL}(V)$ of ω preserving maps:

$$\mathsf{G} := \{A \in \mathsf{GL}(V) \mid \omega(Av, Aw) = g(v, w) \quad \forall v, w \in V\}\,. \tag{186a}$$

Using the "index-lowering" map $\omega^\downarrow : V \to V^*$, $v \mapsto \omega(v, \cdot)$ and its inverse $\omega^\downarrow : V^* \to V$, the "index-raising" map (cf. Sect. A.2), this can also be written as

$$\mathsf{G} := \{A \in \mathsf{GL}(V) \mid \omega^\downarrow \circ A \circ \omega^\uparrow = (A^\top)^{-1}\}\,. \tag{186b}$$

The Lie algebra $\mathrm{Lie}(\mathsf{G})$ is easily obtained by considering curves in G, as explained in the previous subsection. Using (186) this leads to

$$\mathrm{Lie}(\mathsf{G}) := \{X \in \mathrm{End}(V) \mid \omega(Xv, w) + \omega(v, Xw) = 0 \quad \forall v, w \in V\}\,, \tag{187a}$$

$$= \{X \in \mathrm{End}(V) \mid \omega^\downarrow \circ X \circ \omega^\uparrow = -X^\top\}\,. \tag{187b}$$

Let us describe it more concretely in terms of components. Choose a basis $\{e_a\}_{a=1\cdots n}$ of V and the corresponding dual basis $\{\eta^a\}_{a=1\cdots n}$ of V^*, so that $\eta^a(e_b) = \delta^a_b$. From (187b) it follows that a general element $X^a_b\, e_a \otimes \eta^b \in \mathrm{End}(V)$ lies in $\mathrm{Lie}(\mathsf{G})$ iff $X_{ab} = -\,\epsilon\, X_{ba}$, where $X_{ab} := X^c_b \omega_{ca}$. Hence, writing $\eta_a := \omega_{ab}\eta^b$ so that $\eta_a(e_b) = \omega_{ab}$ (cf. Sect. A.5), a basis for $\mathrm{Lie}(\mathsf{G})$ is given by the $\frac{1}{2}n(n-\epsilon)$ vectors

$$M_{ab} = e_a \otimes \eta_b - \,\epsilon\, e_b \otimes \eta_a\,. \tag{188}$$

The Lie algebra of the corresponding inhomogeneous group is given by the linear space $V \oplus \mathrm{Lie}(\mathsf{G})$ and Lie bracket as follows:

$$\big[(a_1, X_1)\,,\, (a_2, X_2)\big] = \big(X_1(a_2) - X_2(a_1)\,,\, [X_1, X_2]\big)\,. \tag{189}$$

Hence we obtain Lie(IG) by adding to (188) the n translation generators

$$T_a := e_a \,. \tag{190}$$

Together they span the $\frac{1}{2}n(n+2-\epsilon)$-dimensional Lie algebra Lie(IG), whose commutation relations easily follow from (188,189,190):

$$[M_{ab}, M_{cd}] \;=\; \omega_{ad}M_{bc} + \omega_{bc}M_{ad} - \epsilon\,\omega_{ac}M_{bd} - \epsilon\,\omega_{bd}M_{ac}\,, \tag{191a}$$

$$[M_{ab}, T_c] \;=\; \omega_{bc}T_a - \epsilon\,\omega_{ac}T_b\,, \tag{191b}$$

$$[T_a, T_b] \;=\; 0\,. \tag{191c}$$

Two special cases of this general setting become relevant in SR:

1 Let $V = \mathbb{R}^4$ and ω symmetric with signature $(1,3)$ (one plus, three minuses). The technical name of G is then O(1,3). Generally, if $V = \mathbb{R}^n$ and if ω is of signature (p,q), where $p+q=n$, G is called O(p,q). O(p,q) is isomorphic to O(q,p) and O(n,0) is just the ordinary orthogonal group O(n) in n dimensions.
2 Let $V = \mathbb{C}^2$ and ω antisymmetric. In two dimensions, leaving an antisymmetric form invariant is equivalent to having unit determinant. Hence G = SL(2,$\mathbb{C}$), the group of complex 2×2 matrices of unit determinant. The group SL(2,$\mathbb{C}$) is the double (and also universal) cover of the identity component of O(1,3), often denoted by $\mathsf{O}^{\uparrow}_{+}(1,3)$ or $\mathsf{SO}^{\uparrow}(1,3)$.

A.10 Exponential Map

Since End(V) form an associative algebra, we can form functions based on addition and multiplication. Writing X^n for the n-fold composition $X\circ\cdots\circ X$, we can define the exponential map

$$\exp : \mathrm{End}(V) \to \mathrm{End}(V)\,, \qquad \exp(X) := \sum_{n=0}^{\infty} \frac{X^n}{n!}\,. \tag{192}$$

Note that the series converges absolutely with respect to the standard norms on End(V).

Now consider "det" and "trace", which are the familiar $\mathbb{F}$-valued functions on End(V):

$$\det(X) \;:=\; \det{}_m\{\eta^a(Ae_b)\}\,, \tag{193}$$

$$\mathrm{trace}(X) \;:=\; \eta^a(Ae_a)\,, \tag{194}$$

where $\{e_a\}$ and $\{\eta^a\}$ is any pair of dual bases (it does not matter which one) and where $\det_m$ is the standard determinant function for matrices. We have

Proposition 11.

$$\det\circ\exp = \exp\circ\,\mathrm{trace}\,. \tag{195}$$

Proof. Assume V to be complex (complexify if V was real). For $X \in \mathrm{End}(V)$ one can then find an eigenbasis, so that with respect to it X is a triangular matrix, whose diagonal entries are its eigenvalues. Then equation (195) reduces to the statement, that the product of the exponentials of the eigenvalues is the exponential of their sum, which is true of course. □

Equation (195) shows that $\det(\exp(X)) > 0$ for any $X \in \mathrm{End}(V)$. Moreover, any element $A = \exp(X)$ is connected to the identity by a continuous path $s \mapsto \exp(sX)$. Hence the image of $\mathrm{End}(V)$ under exp is contained in the identity component of $\mathsf{GL}(V)$, which is given by the invertible linear maps of positive determinant, denoted by $\mathsf{GL}^+(V)$.

Note that the curve $s \mapsto \exp(sX)$ is a homomorphism from the additive group $\mathbb{R}$ to $\mathsf{GL}(V)$. Conversely, we have

Proposition 12. *Let $\gamma : \mathbb{R} \to \mathsf{GL}(V)$ be a homomorphism, i.e. a map that satisfies*

$$\gamma(0) = \mathbf{1} \quad \textit{and} \quad \gamma(s+t) = \gamma(s) \circ \gamma(t) \quad \textit{for all } s, t \in \mathbb{R}. \tag{196}$$

Then γ must be of the form $\gamma(s) = \exp(sX)$, where $X = \dot{\gamma}(0)$.

Proof. We consider the curve $\beta(s) := \gamma(s) \circ \exp(-sX)$, which satisfies $\beta(0) = \mathbf{1}$ and $\dot{\beta}(s) = \dot{\gamma}(s) - \gamma(s) \circ X$. But this is zero, as can be seen from differentiating $\gamma(s+t) = \gamma(s) \circ \gamma(t)$ with respect to t at $t = 0$. Hence $\beta(s) = \mathbf{1}$ for all s, showing that $\gamma(s) = \exp(sX)$. □

Let us now regard the exponential map restricted to the special Lie subalgebras $\mathrm{Lie}(\mathsf{G})$ defined in (187). Since

$$\omega^{\downarrow} \circ \exp(X) \circ \omega^{\uparrow} = \exp(\omega^{\downarrow} \circ X \circ \omega^{\uparrow}) \tag{197}$$

for all $X \in \mathrm{End}(X)$, the image of $\mathrm{Lie}(\mathsf{G})$ under the exponential map lies in G. More precisely, since $\mathrm{Lie}(\mathsf{G})$ is connected and exp continuous, the image must also be connected. Since it also contains the identity ($\mathbf{1} = \exp(0)$), the image of $\mathrm{Lie}(\mathsf{G})$ lies in the identity component of G, denoted by $\mathsf{G}_{\mathbf{1}}$. Hence we have a map

$$\exp : \mathrm{Lie}(\mathsf{G}) \to \mathsf{G}_{\mathbf{1}} \,. \tag{198}$$

It is clear that this map is generally not injective. Consider e.g. the group $\mathsf{SO}(2)$ of planar rotations, which is topologically a circle (S^1) and whose Lie algebra is the real line. exp winds the line infinitely often around the circle. But neither is exp generally onto. A relevant example is given by $\mathsf{G} = \mathsf{SL}(2, \mathbb{C})$. Its Lie algebra is given by the space of traceless $2\times$ matrices with complex entries. Now, for example, none of the matrices

$$A_a := \begin{pmatrix} -1 & a \\ 0 & -1 \end{pmatrix} \quad a \neq 0 \tag{199}$$

can be in the image of the exponential map. To see this, first note that, within $\mathsf{SL}(2, \mathbb{C})$, $A(a)$ can be continuously connected to the identity, e.g. by the path

$$A_a(s) = \begin{pmatrix} \exp(i\pi s) & sa \\ 0 & \exp(-i\pi s) \end{pmatrix} . \tag{200}$$

In fact, $\mathsf{SL}(2,\mathbb{C})$ is connected. Suppose now that $\exp(X) = A_a$ for some traceless X. The eigenvalues of X are $\pm\lambda \neq 0$ so that X is diagonalizable. Let $T \in \mathsf{GL}(2,\mathbb{C})$ such that $TXT^{-1} = \mathrm{diag}(\lambda, -\lambda)$, then $TA_aT^{-1} = \mathrm{diag}(\exp(\lambda), \exp(-\lambda))$, which is impossible since both eigenvalues of A_a equal -1.

What is however true is that the image of the exponential map covers a neighborhood of the group identity. This follows from the fact that the derivative of the smooth map (198) evaluated at $\mathbf{1} \in \mathsf{G}$ is non-zero (it is the identity map $\mathrm{Lie}(\mathsf{G}) \to \mathrm{Lie}(\mathsf{G})$). Hence, by the inverse-function theorem, it has a local smooth inverse. Moreover, we have the following

Proposition 13. *Any $A \in \mathsf{G}$ is the* finite *product of elements in the image of* exp*, that is, for any $A \in \mathsf{G}$ there exist $X_i \in \mathrm{Lie}(\mathsf{G})$, $i = 1, \ldots, k < \infty$, such that*

$$A = \exp(X_1) \circ \cdots \circ \exp(X_k) . \tag{201}$$

Proof. We first note that elements of the form (201) obviously form a subgroup $\mathsf{G}' \subset \mathsf{G}_{\mathbf{1}}$ which contains a whole neighborhood $U \subset \mathbf{1} \in \mathsf{G}_{\mathbf{1}}$, as we have just seen. Now, for any $A \in \mathsf{G}'$, the map $L_A : \mathsf{G}' \to \mathsf{G}'$, $B \mapsto AB$, is a smooth bijection with smooth inverse $L_{A^{-1}}$. Hence L_A is an open map (sends open sets to open sets) so that $L_A(U)$ is an open neighborhood of $A \in \mathsf{G}_{\mathbf{1}}$. This shows that $\mathsf{G}' \subseteq \mathsf{G}_{\mathbf{1}}$ is open. Likewise one shows that all cosets of G' in $\mathsf{G}_{\mathbf{1}}$ are open, since they are obtained as images of G' under L_A for some $A \in \mathsf{G}_{\mathbf{1}}$. But this shows that $\mathsf{G}' \subseteq \mathsf{G}_{\mathbf{1}}$ is also closed, since it is the complement of the union of all G'-cosets different from G' itself. Being open and closed in the connected set $\mathsf{G}_{\mathbf{1}}$, G' is necessarily identical to it. [This argument shows in fact that any neighborhood U of the identity in a topological group generates the identity component, in the sense that any element in the identity component is the finite product of elements from U.] □

Acknowledgements

I sincerely thank Jürgen Ehlers and Claus Lämmerzahl for inviting me to the 2005 Potsdam conference on Special Relativity and also giving me the opportunity to present this material in written form. I am grateful to Matteo Carrera for carefully reading parts of the manuscript and suggesting various improvements.

References

1. Alexander Danilovich Alexandrov. Mappings of spaces with families of cones and space-time transformations. *Annali di Matematica (Bologna)*, 103(8):229–257, 1975.
2. James Anderson. *Principles of Relativity Physics.* Academic Press, New York, 1967.

3. Henri Bacry and Jean-Marc Lévy-Leblond. Possible kinematics. *Journal of Mathematical Physics*, 9(10):1605–1614, 1968.
4. Frank Beckman and Donald Quarles. On isometries of euclidean spaces. *Proceedings of the American Mathematical Society*, 4:810–815, 1953.
5. Enrico Beltrametti and Gianni Cassinelli. *The Logic of Quantum Mechanics*. Encyclopedia of Mathematics and its Application Vol. 15. Addison-Wesley, Reading, Massachusetts, 1981.
6. Marcel Berger. *Geometry*, volume I. Springer Verlag, Berlin, first edition, 1987. Corrected second printing 1994.
7. Marcel Berger. *Geometry*, volume II. Springer Verlag, Berlin, first edition, 1987. Corrected second printing 1996.
8. Vittorio Berzi and Vittorio Gorini. Reciprocity principle and the Lorentz transformations. *Journal of Mathematical Physics*, 10(8):1518–1524, 1969.
9. David Bleecker. *Gauge Theory and Variational Principles*. Number 1 in Global Analysis, Pure and Applied. Addison-Wesley, Reading, Massachusetts, 1981.
10. Hans-Jürgen Borchers and Gerhard Hegerfeld. The structure of space-time transformations. *Communications in Mathematical Physics*, 28:259–266, 1972.
11. Émile Borel. La théorie de la relativité et la cinématique. In *Œuvres de Émile Borel*, volume 3, pages 1809–1811. Editions du Centre National de la Recherche Scientifique, Paris, 1972. First appeared in *Comptes Rendus des séances de l'Académie des Sciences* 156 (1913): 215-217.
12. Max Born. Die Theorie des starren Elektrons in der Kinematik des Relativitätsprinzips. *Annalen der Physik (Leipzig)*, 30:1–56, 1909.
13. Horacio Casini. The logic of causally closed spacetime subsets. *Classical and Quantum Gravity*, 19:6389–6404, 2002.
14. Wojciech Cegła and Arkadiusz Jadczyk. Logics generated by causality structures. covariant representations of the galilei group. *Reports on Mathematical Physics*, 9(3):377–385, 1976.
15. Wojciech Cegła and Arkadiusz Jadczyk. Causal logic of Minkowski space. *Communications in Mathematical Physics*, 57:213–217, 1977.
16. Alexander Chubarev and Iosif Pinelis. Linearity of space-time transformations without the one-to-one, line-onto-line, or constancy-of-speed-of-light assumption. *Communications in Mathematical Physics*, 215:433–441, 2000.
17. Paul Ehrenfest. Gleichförmige Rotation starrer Körper und Relativitätstheorie. *Physikalische Zeitschrift*, 10(23):918, 1909.
18. Vladimir Fock. *The Theory of Space Time and Gravitation*. Pergamon Press, London, first english edition, 1959.
19. Philipp Frank and Hermann Rothe. Über die Transformation der Raumzeitkoordinaten von ruhenden auf bewegte Systeme. *Annalen der Physik (Leipzig)*, 34(5):825–855, 1911.
20. Philipp Frank and Hermann Rothe. Zur Herleitung der Lorentztransformation. *Physikalische Zeitschrift*, 13:750–753, 1912. Erratum: ibid, p. 839.
21. Wilhelm I. Fushchich, Vladimir M. Shtelen, and N.I. Serov. *Symmetry Analysis and Exact Solutions of Equations of Nonlinear Mathematical Physics*. Kluwer Academic Publishers, Dordrecht, 1993.
22. Domenico Giulini. *Advanced Special Relativity*. Oxford University Press, Oxford. To appear.
23. Domenico Giulini. On Galilei invariance in quantum mechanics and the Bargmann superselection rule. *Annals of Physics (New York)*, 249(1):222–235, 1996.
24. Domenico Giulini. Uniqueness of simultaneity. *Britisch Journal for the Philosophy of Science*, 52:651–670, 2001.

25. Domenico Giulini. Das Problem der Trägheit. *Philosophia Naturalis*, 39(2):843–374, 2002.
26. Domenico Giulini and Norbert Straumann. Einstein's impact on the physics of the twentieth century. *Studies in the History and Philisophy of Modern Physics*, to appear, 2006. ArXiv physics/0507107.
27. Rudolf Haag. *Local Quantum Physics*. Springer Verlag, Berlin, first 1991 second revised and enlarged 1996 edition, 1996.
28. Gerhard Hegerfeld. The Lorentz transformations: Derivation of linearity and scale factor. *Il Nuovo Cimento*, 10 A(2):257–267, 1972.
29. Gustav Herglotz. Über den vom Standpunkt des Relativitätsprinzips aus als "starr" zu bezeichnenden Körper. *Annalen der Physik (Leipzig)*, 31:393–415, 1910.
30. Wladimir von Ignatowsky. Einige allgemeine Bemerkungen zum Relativitätsprinzip. *Verhandlungen der Deutschen Physikalischen Gesellschaft*, 12:788–796, 1910.
31. Erdal Inönü and Eugen Wigner. On the cotraction of groups and their representations. *Proceedings of the National Academy of Sciences*, 39(6):510–524, 1953.
32. Nathan Jacobson. *Basic Algebra I*. W.H. Freeman and Co., New York, second edition, 1985.
33. Josef M. Jauch. *Foundations of Quantum Mechanics*. Addison-Wesley, Reading, Massachusetts, 1968.
34. Theodor Kaluza. Zur Relativitätstheorie. *Physikalische Zeitschrift*, 11:977–978, 1910.
35. Felix Klein. *Vergleichende Betrachtungen über neuere geometrische Forschungen*. Verlag von Andreas Deichert, Erlangen, first edition, 1872. Reprinted in *Mathematische Annalen (Leipzig)* 43 (1892) 43-100.
36. Max von Laue. Zur Diskussion über den starren Körper in der Relativitätstheorie. *Physikalische Zeitschrift*, 12:85–87, 1911.
37. Fritz Noether. Zur Kinematik des starren Körpers in der Relativitätstheorie. *Annalen der Physik (Leipzig)*, 31:919–944, 1910.
38. Felix Pirani and Gareth Williams. Rigid motion in a gravitational field. *Séminaire JANET (Mécanique analytique et Mécanique céleste)*, 5e année(8):1–16, 1962.
39. Alfred A. Robb. *Optical Geometry of Motion: A New View of the Theory of Relativity*. W. Heffer & Sons Ltd., Cambridge, 1911.
40. Roman U. Sexl and Helmuth K. Urbantke. *Relativity, Groups, Particles*. Springer Verlag, Wien, first edition, 2001. First english edition, succeeding the 1992 third revised german edition.
41. Arnold Sommerfeld. Über die Zusammensetzung der Geschwindigkeiten in der Relativtheorie. *Physikalische Zeitschrift*, 10:826–829, 1909.
42. Llewellyn Hilleth Thomas. The kinematics of an electron with an axis. *Philosophical Magazine*, 3:1–22, 1927.
43. Andrzej Trautman. Foundations and current problems of general relativity. In A. Trautman, F.A.E. Pirani, and H. Bondi, editors, *Lectures on General Relativity*, volume 1 of *Brandeis Summer Institute in Theoretical Physics*, pages 1–248. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1964.
44. Abraham Ungar. Thomas rotation and the parametrization of the Lorentz transformation group. *Foundations of Physics Letters*, 1(1):57–89, 1988.
45. Abraham Ungar. *Beyond Einstein's Velocity Addition Law*, volume 117 of *Fundamental Theories of Physics*. Kluwer Academic, Dordrecht, 2001.
46. Abraham Ungar. *Analytic Hyperbolic Geometry: Mathematical Foundations and Applications*. World Scientific, Singapore, 2005.

47. Helmuth Urbantke. Physical holonomy: Thomas precession, and Clifford algebra. *American Journal of Physics*, 58(8):747–750, 1990. Erratum ibid. 59(12), 1991, 1150-1151.
48. Helmuth Urbantke. Lorentz transformations from reflections: Some applications. *Foundations of Physics Letters*, 16:111–117, 2003. ArXiv math-ph/0212038.
49. Vladimir Varičak. Anwendung der Lobatschefskijschen Geometrie in der Relativtheorie. *Physikalische Zeitschrift*, 11:93–96, 1910.
50. Vladimir Varičak. Über die nichteuklidische Interpretation der Relativtheorie. *Jahresberichte der Deutschen Mathematikervereinigung (Leipzig)*, 21:103–127, 1912.
51. Erik Christopher Zeeman. Causality implies the Lorentz group. *Journal of Mathematical Physics*, 5(4):490–493, 1964.

Quantum Theory in Accelerated Frames of Reference

B. Mashhoon

Department of Physics and Astronomy, University of Missouri-Columbia
Columbia, Missouri 65211, USA
mashhoonb@missouri.edu

Abstract. The observational basis of quantum theory in accelerated systems is studied. The extension of Lorentz invariance to accelerated systems via the hypothesis of locality is discussed and the limitations of this hypothesis are pointed out. The nonlocal theory of accelerated observers is briefly described. Moreover, the main observational aspects of Dirac's equation in noninertial frames of reference are presented. The Galilean invariance of nonrelativistic quantum mechanics and the mass superselection rule are examined in the light of the invariance of physical laws under inhomogeneous Lorentz transformations.

1 Introduction

Soon after Dirac discovered the relativistic wave equation for a spin $\frac{1}{2}$ particle [1], the generally covariant Dirac equation was introduced by Fock and Ivanenko [2] and was studied in great detail by a number of authors [3]. Dirac's equation

$$(i\hbar\gamma^\alpha\partial_\alpha - mc)\psi = 0 \tag{1}$$

transforms under a Lorentz transformation $x'^\alpha = L^\alpha{}_\beta\, x^\beta$ as

$$\psi'(x') = S(L)\psi(x)\ , \tag{2}$$

where $S(L)$ is connected with the spin of the particle and is given by

$$S^{-1}\gamma^\alpha S = L^\alpha{}_\beta\gamma^\beta. \tag{3}$$

The generally covariant Dirac equation can be written as

$$(i\hbar\gamma^\mu\nabla_\mu - mc)\psi = 0\ , \tag{4}$$

where $\nabla_\mu = \partial_\mu + \Gamma_\mu$ and Γ_μ is the spin connection. Let us consider a class of observers in spacetime with an orthonormal tetrad frame $\lambda^\mu{}_{(\alpha)}$, i.e.

$$g_{\mu\nu}\lambda^\mu{}_{(\alpha)}\lambda^\nu{}_{(\beta)} = \eta_{(\alpha)(\beta)}\ , \tag{5}$$

B. Mashhoon: *Quantum Theory in Accelerated Frames of Reference*, Lect. Notes Phys. **702**, 112–132 (2006)
DOI 10.1007/3-540-34523-X_5 © Springer-Verlag Berlin Heidelberg 2006

where $\eta_{(\alpha)(\beta)}$ is the Minkowski metric tensor. Then in (4), γ^{μ} is given by $\gamma^{\mu} = \lambda^{\mu}{}_{(\alpha)}\gamma^{(\alpha)}$ and

$$\Gamma_{\mu} = -\frac{i}{4}\lambda_{\nu(\alpha)}[\lambda^{\nu}{}_{(\beta)}]_{;\mu}\,\sigma^{(\alpha)(\beta)}\,, \tag{6}$$

where

$$\sigma^{(\alpha)(\beta)} = \frac{i}{2}[\gamma^{(\alpha)}, \gamma^{(\beta)}]\,. \tag{7}$$

In this way, the generally covariant Dirac equation is minimally coupled to inertia and gravitation.

The standard quantum measurement theory involves ideal inertial observers. However, all actual observers are more or less accelerated. Indeed, the whole observational basis of Lorentz invariance as well as quantum mechanics rests upon measurements performed by accelerated observers. It is therefore necessary to discuss how the measurements of noninertial observers are connected with those of ideal inertial observers. This paper is thus organized into two parts. In the first part, Sects. 2–4, we consider the basic physical assumptions that underlie the covariant generalization of Dirac's equation. The second part, Sects. 5–9, are devoted to the physical consequences of this generalization for noninertial frames of reference. In particular, the connection between the relativistic theory and nonrelativistic quantum mechanics in accelerated systems is examined in detail. Sect. 10 contains a brief discussion.

2 Hypothesis of Locality

The extension of Lorentz invariance to noninertial systems necessarily involves an assumption regarding what accelerated observers actually measure. What is assumed in the standard theory of relativity is the *hypothesis of locality,* which states that an accelerated observer is pointwise equivalent to an otherwise identical momentarily comoving inertial observer. It appears that Lorentz first introduced such an assumption in his theory of electrons to ensure that an electron – conceived as a small ball of charge – is always Lorentz contracted along its direction of motion [4]. He clearly recognized that this is simply an approximation based on the assumption that the time in which the electron velocity changes is very long compared to the period of the internal oscillations of the electron (see Sect. 183 on page 216 of [4]).

The hypothesis of locality was later adopted by Einstein in the course of the development of the theory of relativity (see the footnote on page 60 of [5]). In retrospect, the locality assumption fits perfectly together with Einstein's local principle of equivalence to guarantee that every observer in a gravitational field is locally (i.e. pointwise) inertial. That is, Einstein's heuristic principle of equivalence, namely, the presumed local equivalence of an observer in a gravitational field with an accelerated observer in Minkowski spacetime, would lose its operational significance if one did not know what accelerated observers measure.

However, combined with the hypothesis of locality, Einstein's principle of equivalence provides a basis for a theory of gravitation that is consistent with (local) Lorentz invariance.

Early in the development of the theory of relativity, the hypothesis of locality was usually stated in terms of the direct acceleration independence of the behavior of rods and clocks. The clock hypothesis, for instance, states that "standard" clocks measure proper time. Thus measuring devices that conform to the hypothesis of locality are usually called "standard". It is clear that inertial effects exist in any accelerated measuring device; however, in a *standard* device these effects are usually expected to integrate to a negligible influence over the duration of each elementary measurement. Thus a standard measuring device is locally inertial [6].

Following the development of the general theory of relativity, the hypothesis of locality was discussed by Weyl [7]. Specifically, Weyl [7] noted that the locality hypothesis was an adiabaticity assumption in analogy with slow processes in thermodynamics.

The hypothesis of locality originates from Newtonian mechanics: the accelerated observer and the otherwise identical momentarily comoving inertial observer have the same position and velocity; therefore, they share the same *state* and are thus pointwise identical in classical mechanics. The evident validity of this assertion for Newtonian point particles means that no new assumption is required in the treatment of accelerated systems of reference in Newtonian mechanics. It should also hold equally well in the classical *relativistic* mechanics of point particles, as originally recognized by Minkowski (see p. 80 of [8]). If all physical phenomena could be reduced to *pointlike coincidences* of particles and rays, then the hypothesis of locality would be exactly valid.

The hypothesis of locality is not in general valid, however, in the case of classical wave phenomena. Consider, for instance, the determination of the frequency of an incident electromagnetic wave by a linearly accelerated observer. Clearly, the frequency cannot be determined instantaneously; in fact, the observer needs to measure a few oscillations of the electromagnetic field before a reasonable determination of the frequency becomes operationally possible. Let λ be the characteristic wavelength of the incident radiation and $\mathcal{L}$ be the acceleration length of the observer; then, the hypothesis of locality is approximately valid for $\lambda \ll \mathcal{L}$. Here $\mathcal{L}$ is a length scale that involves the speed of light c and certain scalars formed from the acceleration of the observer such that the acceleration time $\mathcal{L}/c$ characterizes the time in which the velocity of the observer varies appreciably. In an Earth-based laboratory, for instance, the main translational and rotational acceleration lengths would be $c^2/g_\oplus \approx 1$ lt-yr and $c/\Omega_\oplus \approx 28\,\mathrm{AU}$, respectively. Thus in most experimental situations $\lambda/\mathcal{L}$ is negligibly small and any possible deviations from the locality hypothesis are therefore below the current levels of detectability. Indeed, in the ray limit, $\lambda/\mathcal{L} \to 0$, the hypothesis of locality would be valid; therefore, $\lambda/\mathcal{L}$ is a measure of possible deviation from the locality postulate.

Consider a classical particle of mass m and charge q under the influence of an external force $\mathbf{f}_{\rm ext}$. The accelerated charge radiates electromagnetic radiation with a typical wavelength $\lambda \sim \mathcal{L}$, where $\mathcal{L}$ is the acceleration length of the particle.We would expect that a significant breakdown of the locality hypothesis occurs in this case, since $\lambda\!\!\!^{-}/\mathcal{L} \sim 1$ in the interaction of the particle with the electromagnetic field. The violation of the hypothesis of locality implies that the state of the particle cannot be characterized by its position and velocity. This is indeed the case, since the equation of motion of the radiating particle in the nonrelativistic approximation is given by the Abraham-Lorentz equation

$$m\frac{d\mathbf{v}}{dt} - \frac{2}{3}\frac{q^2}{c^3}\frac{d^2\mathbf{v}}{dt^2} + \cdots = \mathbf{f}_{\rm ext} \tag{8}$$

which implies that position and velocity are not sufficient to specify the state of the radiating charged particle [9].

To discuss quantum mechanics in an accelerated system of reference, it is therefore useful to investigate the status of the hypothesis of locality vis-a-vis the basic principles of quantum theory. The physical interpretation of wave functions is based on the notion of wave-particle duality. On the other hand, the locality hypothesis is valid for classical particles and is in general violated for classical waves. This circumstance provides the motivation to develop a nonlocal theory of accelerated systems that would go beyond the hypothesis of locality and would be consistent with wave-particle duality. Such a theory has been developed [10] and can be employed, in principle, to describe a nonlocal Dirac equation in accelerated systems of reference. Some of the main aspects of the nonlocal theory are described in Sect. 4.

3 Acceleration Tensor

It follows from the hypothesis of locality that an accelerated observer in Minkowski spacetime carries an orthonormal tetrad $\lambda^{\mu}{}_{(\alpha)}$, where $\lambda^{\mu}{}_{(0)} = dx^{\mu}/d\tau$ is its four-velocity vector that is tangent to its worldline and acts as its local temporal axis. Here τ is the proper time along the worldline of the accelerated observer. To avoid unphysical situations, we assume throughout that the observer is accelerated only for a finite period of time. The local spatial frame of the observer is defined by the unit spacelike axes $\lambda^{\mu}{}_{(i)}$, $i = 1, 2, 3$. The tetrad frame is transported along the worldline in accordance with

$$\frac{d\lambda^{\mu}{}_{(\alpha)}}{d\tau} = \Phi_{\alpha}{}^{\beta}\lambda^{\mu}{}_{(\beta)} \ , \tag{9}$$

where

$$\Phi_{\alpha\beta} = -\Phi_{\beta\alpha} \tag{10}$$

is the antisymmetric *acceleration tensor*. In close analogy with the Faraday tensor, the acceleration tensor consists of "electric" and "magnetic" components.

The "electric" part is characterized by the translational acceleration of the observer such that $\Phi_{0i} = a_i(\tau)$, where $a_i = A_\mu \lambda^\mu{}_{(i)}$ and $A^\mu = d\lambda^\mu{}_{(0)}/d\tau$ is the four-acceleration vector of the observer. The "magnetic" part is characterized by the rotation of the local spatial frame with respect to a locally nonrotating (i.e. Fermi-Walker transported) frame such that $\Phi_{ij} = \epsilon_{ijk}\Omega^k$, where $\boldsymbol{\Omega}(\tau)$ is the rotation frequency. The elements of the acceleration tensor, and hence the spacetime scalars $\mathbf{a}(\tau)$ and $\boldsymbol{\Omega}(\tau)$, completely determine the local rate of variation of the state of the observer. It proves useful to define the acceleration lengths $\mathcal{L} = c^2/a$ and c/Ω, as well as the corresponding acceleration times $\mathcal{L}/c = c/a$ and $1/\Omega$, to indicate respectively the spatial and temporal scales of variation of the state of the observer. Let λ be the intrinsic length scale of the phenomenon under observation; then, we expect that the deviation from the hypothesis of locality should be proportional to $\lambda/\mathcal{L}$.

It follows from a detailed analysis that if D is the spatial dimension of a standard measuring device, then $D \ll \mathcal{L}$ [6]. Such devices are necessary for the determination of the local frame of the accelerated observer. In fact, this circumstance is analogous to the correspondence principle: while we are interested in the deviations from the hypothesis of locality, such nonlocal effects are expected to be measured with standard measuring devices.

4 Nonlocality

Imagine an accelerated observer in a background global Minkowski spacetime and let $\psi(x)$ be a basic incident radiation field. The observer along its worldline passes through a continuous infinity of hypothetical momentarily comoving inertial observers; therefore, let $\hat{\psi}(\tau)$ be the field measured by the hypothetical inertial observer at the event characterized by the proper time τ. The local spacetime of the hypothetical inertial observer is related to the background via a proper Poincaré transformation $x' = Lx + s$; hence, $\psi'(x') = \Lambda(L)\psi(x)$, so that $\Lambda = 1$ for a scalar field. We therefore assume that along the worldline $\hat{\psi}(\tau) = \Lambda(\tau)\psi(\tau)$, where Λ belongs to a matrix representation of the Lorentz group.

Suppose that $\hat{\Psi}(\tau)$ is the field that is actually measured by the accelerated observer. What is the connection between $\hat{\Psi}(\tau)$ and $\hat{\psi}(\tau)$? The hypothesis of locality postulates the pointwise equivalence of $\hat{\Psi}(\tau)$ and $\hat{\psi}(\tau)$, i.e. it requires that $\hat{\Psi}(\tau) = \hat{\psi}(\tau)$. On the other hand, the most general linear relation between $\hat{\Psi}(\tau)$ and $\hat{\psi}(\tau)$ consistent with causality is

$$\hat{\Psi}(\tau) = \hat{\psi}(\tau) + \int_{\tau_0}^{\tau} K(\tau, \tau')\hat{\psi}(\tau')d\tau' \,, \tag{11}$$

where τ_0 is the initial instant of the observer's acceleration. Equation (11) is manifestly Lorentz invariant, since it involves spacetime scalars. The kernel $K(\tau, \tau')$ must be directly proportional to the observer's acceleration, since $\hat{\Psi} = \hat{\psi}$ for an inertial observer. The ansatz (11) differs from the hypothesis of locality by

an integral over the past worldline of the observer. In fact, this nonlocal part is expected to vanish for $\lambda/\mathcal{L} \to 0$. The determination of a radiation field by an accelerated observer involves a certain spacetime average according to (11) and this circumstance is consistent with the viewpoint developed by Bohr and Rosenfeld [11].

Equation (11) has the form of a Volterra integral equation. According to Volterra's theorem [12], the relationship between $\hat{\Psi}$ and $\hat{\psi}$ (and hence ψ) is unique in the space of continuous functions. Volterra's theorem has been extended to the Hilbert space of square-integrable functions by Tricomi [13].

To determine the kernel K, we postulate that a basic radiation field can never stand completely still with respect to an accelerated observer. This physical requirement is a generalization of a well-known consequence of Lorentz invariance to all observers. That is, the invariance of Maxwell's equations under the Lorentz transformations implies that electromagnetic radiation propagates with speed c with respect to all inertial observers. That this is the case for any basic radiation field is reflected in the Doppler formula, $\omega' = \gamma(\omega - \mathbf{v} \cdot \mathbf{k})$, where $\omega = c|\mathbf{k}|$. An inertial observer moving uniformly with speed v that approaches c measures a frequency ω' that approaches zero, but the wave will never stand completely still ($\omega' \neq 0$) since $v < c$; hence, $\omega' = 0$ implies that $\omega = 0$. Generalizing this situation to arbitrary accelerated observers, we demand that if $\hat{\Psi}$ turns out to be a constant, then ψ must have been constant in the first place. The Volterra-Tricomi uniqueness result then implies that for any true radiation field ψ in the inertial frame, the field $\hat{\Psi}$ measured by the accelerated observer will vary in time. Writing (11) as

$$\hat{\Psi}(\tau) = \Lambda(\tau)\psi(\tau) + \int_{\tau_0}^{\tau} K(\tau, \tau')\Lambda(\tau')\psi(\tau')d\tau' \,, \tag{12}$$

we note that our basic postulate that a constant $\hat{\Psi}$ be associated with a constant ψ implies

$$\Lambda(\tau_0) = \Lambda(\tau) + \int_{\tau_0}^{\tau} K(\tau, \tau')\Lambda(\tau')d\tau' \,, \tag{13}$$

where we have used the fact that $\hat{\Psi}(\tau_0) = \Lambda(\tau_0)\psi(\tau_0)$. Given $\Lambda(\tau)$, (13) can be used to determine $K(\tau, \tau')$; however, it turns out that $K(\tau, \tau')$ cannot be uniquely specified in this way. To go forward, it originally appeared most natural from the standpoint of phenomenological nonlocal theories to postulate that $K(\tau, \tau')$ is only a function of $\tau - \tau'$ [10]; however, detailed investigations later revealed that such a convolution kernel can lead to divergences in the case of nonuniform acceleration [14]. It turns out that the only physically acceptable solution of (13) is of the form [15, 16]

$$K(\tau, \tau') = k(\tau') = -\frac{d\Lambda(\tau')}{d\tau'}\Lambda^{-1}(\tau') \,. \tag{14}$$

In the case of uniform acceleration, (14) and the convolution kernel both lead to the same constant kernel. The kernel (14) is directly proportional to the

acceleration of the observer and is a simple solution of (13), as can be verified by direct substitution. Moreover, if the acceleration of the observer is turned off at τ_f, then the unique kernel (14) vanishes for $\tau > \tau_f$. Thus for $\tau > \tau_f$, the nonlocal contribution to the field in (11) is simply a constant memory of the past acceleration of the observer that is in principle measurable. This constant memory is simply canceled in a measuring device whenever the device is reset.

For a scalar field $\Lambda = 1$ and hence the kernel (14) vanishes. As will be demonstrated in Sect. 8, it follows from the locality of such a field that for scalar radiation of frequency ω, an observer rotating uniformly with frequency Ω will measure $\omega' = \gamma(\omega - M\Omega)$, where $M = 0, \pm 1, \pm 2, \ldots$. Thus $\omega' = 0$ for $\omega = M\Omega$ and our basic physical postulate is violated: the scalar radiation stands completely still for all observers rotating uniformly about the same axis with frequency Ω. It therefore follows from the nonlocal theory of accelerated observers that a pure scalar (or pseudoscalar) radiation field does not exist. Such fields can only be composites formed from other basic fields. This consequence of the nonlocal theory is consistent with present observational data, as they show no trace of a fundamental scalar (or pseudoscalar) field.

4.1 Nonlocal Field Equations

It follows from the Volterra (11) with kernel (14) that

$$\hat{\psi} = \hat{\Psi} + \int_{\tau_0}^{\tau} r(\tau, \tau') \hat{\Psi}(\tau') d\tau' \,, \tag{15}$$

where $r(\tau, \tau')$ is the resolvent kernel. Imagine that a nonlocal field Ψ exists in the background Minkowski spacetime such that an accelerated observer with a tetrad frame $\lambda^{\mu}{}_{(\alpha)}$ measures

$$\hat{\Psi} = \Lambda \Psi \,. \tag{16}$$

The relationship between Ψ and ψ can then be simply worked out using (15), namely,

$$\psi = \Psi + \int_{\tau_0}^{\tau} \tilde{r}(\tau, \tau') \Psi(\tau') d\tau' \,, \tag{17}$$

where $\tilde{r}$ is related to the resolvent kernel by

$$\tilde{r}(\tau, \tau') = \Lambda^{-1}(\tau) r(\tau, \tau') \Lambda(\tau') \,. \tag{18}$$

It is possible to extend (17) to a class of accelerated observers such that $\psi(x)$ within a finite region of spacetime is related to a nonlocal field $\Psi(x)$ by a suitable extension of (17). The local field $\psi(x)$ satisfies certain partial differential equations; therefore, it follows from (17) that Ψ would satisfy certain Lorentz-invariant nonlocal field equations. In this way, the nonlocal Maxwell equations have been derived explicitly for certain linearly accelerated systems [17]. It turns out that in general the field equations remain nonlocal even after the cessation of accelerated motion.

4.2 Nonlocal Electrodynamics

To confront the nonlocal theory with observation, it is useful to derive the physical consequences of nonlocal electrodynamics in systems that undergo translational and rotational accelerations and compare the predictions of the theory with observational data. It turns out that for accelerated systems the experimental data available at present do not have sufficient sensitivity to distinguish between the standard theory (based on the locality hypothesis) and the nonlocal theory. In the case of linearly accelerated systems, it may be possible to reach the desired level of sensitivity with the acceleration of grains using high-intensity femtosecond lasers [18, 19]. For a uniformly rotating observer in circular motion, one can compare the predictions of nonlocal electrodynamics with the nonrelativistic quantum mechanics of electrons in circular atomic orbits or about uniform magnetic fields in the correspondence limit. If the nonlocal theory corresponds to reality, its predictions should be closer to quantum mechanical results in the correspondence regime than those of the standard local theory of accelerated systems. This turns out to be the case for the simple cases that have been worked out in detail [20]. Let us now return to the standard physical consequences of Dirac's equation in noninertial systems of reference. In the following sections, emphasis will be placed on the main inertial effects and their observational aspects in matter-wave interferometry.

5 Inertial Properties of a Dirac Particle

The physical consequences that follow from the Dirac equation in systems of reference that undergo translational and rotational accelerations have been considered by a number of authors [21–24]. In particular, the work of Hehl and Ni [25] has elucidated the general inertial properties of a Dirac particle. In their approach, standard Foldy-Wouthuysen [26] transformations are employed to decouple the positive and negative energy states such that the Hamiltonian for the Dirac particle may be written as

$$\mathcal{H} = \beta\left(mc^2 + \frac{p^2}{2m}\right) + \beta m\boldsymbol{a}\cdot\boldsymbol{x} - \boldsymbol{\Omega}\cdot(\boldsymbol{L}+\boldsymbol{S}) \tag{19}$$

plus higher-order terms. Here $\beta m\boldsymbol{a}\cdot\boldsymbol{x}$ is an inertial term due to the translational acceleration of the reference frame, while the inertial effects due to the rotation of the reference frame are reflected in $-\boldsymbol{\Omega}\cdot(\boldsymbol{L}+\boldsymbol{S})$.

Before proceeding to a detailed discussion of these inertial terms in Sects. 6–9, it is important to observe that Obukhov [27] has recently introduced certain exact "Foldy-Wouthuysen" (FW) transformations to decouple the positive and negative energy states of the Dirac particle. Such a FW transformation is defined up to a unitary transformation, which introduces a certain level of ambiguity in the physical interpretation. That is, it is not clear from [27] what one could predict to be the observable consequences of Dirac's theory in noninertial systems

and gravitational fields. For instance, in Obukhov's exact FW transformation, an inertial term of the form $-\frac{1}{2}\boldsymbol{S}\cdot\boldsymbol{a}$ appears in the Hamiltonian [27]; on the other hand, it is possible to remove this term by a unitary transformation [27]. The analog of this term in a gravitational context would be $\frac{1}{2}\boldsymbol{S}\cdot\boldsymbol{g}$. Thus the energy difference between the states of a Dirac particle with spin polarized up and down in a laboratory on the Earth would be $\frac{1}{2}\hbar g_{\oplus} \approx 10^{-23}\,\mathrm{eV}$, which is a factor of five larger than what can be detected at present [28]. A detailed examination of spin-acceleration coupling together with theoretical arguments for its absence is contained in [29].

The general question raised in [27] has been treated in [30]. It appears that with a proper choice of the unitary transformation such that physical quantities would correspond to simple operators, the standard FW transformations of Hehl and Ni [25] can be recovered [30]. Nevertheless, a certain phase ambiguity can still exist in the wave function corresponding to the fact that the unitary transformation may not be unique. This phase problem exists even in the nonrelativistic treatment of quantum mechanics in translationally accelerated systems as discussed in detail in Sect. 9.

6 Rotation

It is possible to provide a simple justification for the rotational inertial term in the Hamiltonian (19). Let us start with the classical nonrelativistic Lagrangian of a particle $L = \frac{1}{2}mv^2 - W$, where W is a potential energy. Under a transformation to a rotating frame of reference, $\boldsymbol{v} = \boldsymbol{v}' + \boldsymbol{\Omega}\times\boldsymbol{r}$, the Lagrangian takes the form

$$L' = \frac{1}{2}m(\boldsymbol{v}' + \boldsymbol{\Omega}\times\boldsymbol{r})^2 - W\ , \tag{20}$$

where W is assumed to be invariant under the transformation to the rotating frame. The canonical momentum of the particle $\boldsymbol{p}' = \partial L'/\partial\boldsymbol{v}' = \boldsymbol{p}$ is an invariant and we find that $H' = H - \boldsymbol{\Omega}\cdot\boldsymbol{L}$, where $\boldsymbol{L} = \boldsymbol{r}\times\boldsymbol{p}$ is the invariant angular momentum of the particle. Let us note that this result of Newtonian mechanics [31] has a simple relativistic generalization: the rotating observer measures the energy of the particle to be $E' = \gamma(E - \boldsymbol{v}\cdot\boldsymbol{p})$, where $\boldsymbol{v} = \boldsymbol{\Omega}\times\boldsymbol{r}$; therefore, $E' = \gamma(E - \boldsymbol{\Omega}\cdot\boldsymbol{L})$.

This local approach may be simply extended to nonrelativistic quantum mechanics, where the hypothesis of locality would imply that [32]

$$\psi'(\boldsymbol{x}', t) = \psi(\boldsymbol{x}, t)\ , \tag{21}$$

since the rotating measuring devices are assumed to be locally inertial. Thus $\psi'(\boldsymbol{x}', t) = R\psi(\boldsymbol{x}', t)$, where

$$R = \hat{T}e^{\frac{i}{\hbar}\int_0^t \boldsymbol{\Omega}(t')\cdot\boldsymbol{J}dt'}\ . \tag{22}$$

Here $\hat{T}$ is the time-ordering operator and we have replaced $\boldsymbol{L}$ by $\boldsymbol{J} = \boldsymbol{L} + \boldsymbol{S}$, since the *total* angular momentum is the generator of rotations [32]. It follows

that from the standpoint of rotating observers, $H\psi = i\hbar\partial\psi/\partial t$ takes the form $H'\psi' = i\hbar\partial\psi'/\partial t$, where

$$H' = RHR^{-1} - \boldsymbol{\Omega}\cdot\boldsymbol{J}\ . \tag{23}$$

For the case of the single particle viewed by uniformly rotating observers, H' can be written as

$$H' = \frac{1}{2m}(\boldsymbol{p}' - m\boldsymbol{\Omega}\times\boldsymbol{r})^2 - \frac{1}{2}m(\boldsymbol{\Omega}\times\boldsymbol{r})^2 - \boldsymbol{\Omega}\cdot\boldsymbol{S} + W\ , \tag{24}$$

where $-\frac{1}{2}m(\boldsymbol{\Omega}\times\boldsymbol{r})^2$ is the standard centrifugal potential and $-\boldsymbol{\Omega}\cdot\boldsymbol{S}$ is the spin-rotation coupling term [32]. The Hamiltonian (24) is analogous to that of a charged particle in a uniform magnetic field; this situation is a reflection of the Larmor theorem. The corresponding analog of the Aharonov-Bohm effect is the Sagnac effect for matter waves [33]. This effect is discussed in the next section.

7 Sagnac Effect

The term $-\boldsymbol{\Omega}\cdot\boldsymbol{L}$ in the Hamiltonian (19) signifies the coupling of the orbital angular momentum of the particle with the rotation of the reference frame and is responsible for the Sagnac effect exhibited by the Dirac particle. The corresponding Sagnac phase shift is given by

$$\Delta\Phi_{\text{Sagnac}} = \frac{2m}{\hbar}\int \boldsymbol{\Omega}\cdot d\boldsymbol{A}\ , \tag{25}$$

where $\boldsymbol{A}$ is the area of the interferometer. Equation (25) can be expressed as

$$\Delta\Phi_{\text{Sagnac}} = \frac{2\omega}{c^2}\int \boldsymbol{\Omega}\cdot d\boldsymbol{A}\ , \tag{26}$$

where $mc^2 \approx \hbar\omega$ and ω is the de Broglie frequency of the particle. Equation (26) is equally valid for electromagnetic radiation of frequency ω.

For matter waves, the Sagnac effect was first experimentally measured for Cooper pairs in a rotating superconducting Josephson-junction interferometer [34]. Using slow neutrons, Werner et al. [35] measured the Sagnac effect with $\boldsymbol{\Omega}$ as the rotation frequency of the Earth. The result was subsequently confirmed with a rotating neutron interferometer in the laboratory [36]. Significant advances in atom interferometry have led to the measurement of the Sagnac effect for neutral atoms as well. This was first achieved by Riehle et al. [37] and has been subsequently developed with a view towards achieving high sensitivity for atom interferometers as inertial sensors [38]. In connection with charged particle interferometry, the Sagnac effect has been observed for electrons by Hasselbach and Nicklaus [39].

The Sagnac effect has significant and wide-ranging applications and has been reviewed in [40].

8 Spin-Rotation Coupling

The transformation of the wave function to a uniformly rotating system of coordinates involves $(t, r, \theta, \phi) \to (t, r, \theta, \phi + \Omega t)$ in spherical coordinates, where Ω is the frequency of rotation about the z axis. If the dependence of the wave function on ϕ and t is of the form $\exp(iM\phi - iEt/\hbar)$, then in the rotating system the temporal dependence of the wave function is given by $\exp[-i(E - \hbar M\Omega)t/\hbar]$. The energy of the particle measured by an observer at rest in the rotating frame is

$$E' = \gamma(E - \hbar M\Omega) \,, \tag{27}$$

where $\gamma = t/\tau$ is the Lorentz factor due to time dilation. Here $\hbar M$ is the *total* angular momentum of the particle along the axis of rotation; in fact, $M = 0, \pm 1, \pm 2, \ldots$, for a scalar or a vector particle, while $M \mp \frac{1}{2} = 0, \pm 1, \pm 2, \ldots$, for a Dirac particle.

In the JWKB approximation, (27) may be expressed as $E' = \gamma(E - \boldsymbol{\Omega} \cdot \boldsymbol{J})$ and hence

$$E' = \gamma(E - \boldsymbol{\Omega} \cdot \boldsymbol{L}) - \gamma \boldsymbol{\Omega} \cdot \boldsymbol{S} \,. \tag{28}$$

It follows that the energy measured by the observer is the result of an instantaneous Lorentz transformation together with an additional term

$$\delta H = -\gamma \boldsymbol{\Omega} \cdot \boldsymbol{S} \,, \tag{29}$$

which is due to the coupling of the intrinsic spin of the particle with the frequncy of rotation of the observer [32]. The dynamical origin of this term can be simply understood on the basis of the following consideration: The intrinsic spin of a free particle remains fixed with respect to the underlying global inertial frame; therefore, from the standpoint of observers at rest in the rotating system, the spin precesses in the opposite sense as the rotation of the observers. The Hamiltonian responsible for this inertial motion is given by (29). The relativistic nature of spin-rotation coupling has been demonstrated by Ryder [41]. Let us illustrate these ideas by a thought experiment involving the reception of electromagnetic radiation of frequency ω by an observer that rotates uniformly with frequency Ω. We assume for the sake of simplicity that the plane circularly polarized radiation is normally incident on the path of the observer, i.e. the wave propagates along the axis of rotation. We are interested in the frequency of the wave ω' as measured by the rotating observer. A simple application of the hypothesis of locality leads to the conclusion that the measured frequency is related to ω by the transverse Doppler effect, $\omega'_D = \gamma\omega$, since the instantaneous rest frame of the observer is related to the background global inertial frame by a Lorentz transformation. On the other hand, a different answer emerges when we focus attention on the measured electromagnetic field rather than the propagation vector of the radiation,

$$F_{(\alpha)(\beta)}(\tau) = F_{\mu\nu}\lambda^{\mu}{}_{(\alpha)}\lambda^{\nu}{}_{(\beta)} \,, \tag{30}$$

where $F_{\mu\nu}$ is the Faraday tensor of the incident radiation and $\lambda^{\mu}{}_{(\alpha)}$ is the orthonormal tetrad of the rotating observer. The nonlocal process of Fourier analysis of $F_{(\alpha)(\beta)}$ results in [42]

$$\omega' = \gamma(\omega \mp \Omega) \ , \tag{31}$$

where the upper (lower) sign refers to positive (negative) helicity radiation. We note that in the eikonal limit $\Omega/\omega \to 0$ and the instantaneous Doppler result is recovered. The general problem of electromagnetic waves in a (uniformly) rotating frame of reference has been treated in [43].

It is possible to understand (31) in terms of the relative motion of the observer with respect to the field. In a positive (negative) helicity wave, the electric and magnetic fields rotate with the wave frequency ω $(-\omega)$ about its direction of propagation. Thus the rotating observer perceives that the electric and magnetic fields rotate with frequency $\omega - \Omega$ $(-\omega - \Omega)$ about the direction of wave propagation. Taking due account of time dilation, the observed frequency of the wave is thus $\gamma(\omega - \Omega)$ in the positive helicity case and $\gamma(\omega + \Omega)$ in the negative helicity case. These results illustrate the phenomenon of helicity-rotation coupling for the photon, since (31) can be written as $E' = \gamma(E - \boldsymbol{S} \cdot \boldsymbol{\Omega})$, where $E = \hbar\omega$, $\boldsymbol{S} = \hbar\hat{\boldsymbol{H}}$ and $\hat{\boldsymbol{H}} = \pm c\boldsymbol{k}/\omega$ is the unit helicity vector.

It follows from (31) that for a slowly moving detector $\gamma \approx 1$ and

$$\omega' \approx \omega \mp \Omega \ , \tag{32}$$

which corresponds to the phenomenon of *phase wrap-up* in the Global Positioning System (GPS) [44]. In fact, (32) has been verified for $\omega/(2\pi) \approx 1$ GHz and $\Omega/(2\pi) \approx 8$ Hz by means of the GPS [44]. For $\omega \gg \Omega$, the modified Doppler and aberration formulas due to the helicity-rotation coupling are [45]

$$\omega' = \gamma[(\omega - \hat{\boldsymbol{H}} \cdot \boldsymbol{\Omega}) - \boldsymbol{v} \cdot \boldsymbol{k}] \ , \tag{33}$$

$$\boldsymbol{k}' = \boldsymbol{k} + \frac{1}{v^2}(\gamma - 1)(\boldsymbol{v} \cdot \boldsymbol{k})\boldsymbol{v} - \frac{1}{c^2}\gamma(\omega - \hat{\boldsymbol{H}} \cdot \boldsymbol{\Omega})\boldsymbol{v} \ , \tag{34}$$

and similar formulas can be derived for any spinning particle. Circularly polarized radiation is routinely employed for radio communication with artificial satellites as well as Doppler tracking of spacecraft. In general, the rotation of the emitter as well as the receiver should be taken into account. It follows from (33) that ignoring helicity-rotation coupling would lead to a systematic Doppler bias of magnitude $c\Omega/\omega$. In the case of the Pioneer spacecraft, the anomalous acceleration resulting from the helicity-rotation coupling has been shown to be negligibly small [46].

A half-wave plate flips the helicity of a photon that passes through it. Imagine a half-wave plate that rotates uniformly with frequency Ω and an incident positive helicity plane wave of frequency ω_{in} that propagates along the axis of rotation. It follows from (32) that $\omega' \approx \omega_{\text{in}} - \Omega$. The spacetime of a uniformly rotating system is stationary; therefore, ω' remains fixed inside the plate. The radiation that emerges from the plate has frequency ω_{out} and negative helicity; hence, (32) implies that $\omega' \approx \omega_{\text{oout}} + \Omega$. Thus the rotating half-wave plate is a frequency shifter: $\omega_{\text{out}} - \omega_{\text{in}} \approx -2\Omega$. In general, any rotating spin flipper can cause an up/down energy shift given by $-2\boldsymbol{S} \cdot \boldsymbol{\Omega}$ as a consequence of the spin-rotation coupling. The frequency-shift phenomenon was first discovered in

microwave experiments [47] and has subsequently been used in many optical experiments (see [45] for a list of references).

Regarding the spin-rotation coupling for fermions, let us note that for experiments in a laboratory fixed on the Earth, we must add to every Hamiltonian the spin-rotation-gravity term

$$\delta H \approx -\boldsymbol{S}\cdot\boldsymbol{\Omega} + \boldsymbol{S}\cdot\boldsymbol{\Omega}_P \,, \tag{35}$$

where the second term is due to the gravitomagnetic field of the Earth. That is, the rotation of the Earth causes a dipolar gravitomagnetic field (due to mass current), which is locally equivalent to a rotation by the gravitational Larmor theorem. In fact, $\boldsymbol{\Omega}_P$ is the frequency of precession of an ideal fixed test gyro and is given by

$$\boldsymbol{\Omega}_P \approx \frac{G}{c^2 r^5}[3(\boldsymbol{J}\cdot\boldsymbol{r})\boldsymbol{r} - \boldsymbol{J}r^2] \,, \tag{36}$$

where J is the proper angular momentum of the central source. It follows from (35) that for a spin $\frac{1}{2}$ particle, the difference between the energy of the particle with spin up and down in the laboratory is characterized by $\hbar\Omega_\oplus \sim 10^{-19}\,\mathrm{eV}$ and $\hbar\Omega_P \sim 10^{-29}\,\mathrm{eV}$, while the present experimental capabilities are in the $10^{-24}\,\mathrm{eV}$ range [28]. In fact, indirect observational evidence for the spin-rotation coupling has been obtained [48] from the analysis of experiments that have searched for anomalous spin-gravity interactions [49]. Further evidence for spin-rotation coupling exists based on the analysis of muon $g-2$ experiment [51].

An experiment to measure directly the spin-rotation coupling for a spin $\frac{1}{2}$ particle was originally proposed in [32]. This involved a large-scale neutron interferometry experiment with polarized neutrons on a rotating platform [52]. A more recent proposal [53] employs a rotating neutron spin flipper and hence is much more manageable as it avoids a large-scale interferometer. The slow neutrons from a source are longitudinally polarized and the beam is coherently split into two paths that contain neutron spin flippers, one of which rotates with frequency Ω about the direction of motion of the neutrons. In this leg of the interferometer, an energy shift $\delta H = -2\boldsymbol{S}\cdot\boldsymbol{\Omega}$ is thus introduced. The two beams are brought back together and the interference beat frequency Ω is then measured. It is interesting to note that a beat frequency in neutron interferometry has already been measured in another context [54]; therefore, similar techniques can be used in the proposed experiment [53].

Some general remarks on the calculation of the phase shift are in order here. One starts from the relation $\hbar\, d\Phi = -E dt + \boldsymbol{p}\cdot d\boldsymbol{x}$ for the phase $\Phi(\boldsymbol{x},t)$ of the neutron wave in the JWKB approximation. Integrating from the source $(\boldsymbol{x}_S, t_S)$ to the detector $(\boldsymbol{x}_D, t_D)$, we find

$$\hbar\Phi(\boldsymbol{x}_D,t_D) = \hbar\Phi(\boldsymbol{x}_S,t_S) - \int_{t_S}^{t_D} E dt + \int_{\boldsymbol{x}_S}^{\boldsymbol{x}_D} \boldsymbol{p}\cdot d\boldsymbol{x} \,. \tag{37}$$

Assuming equal amplitudes, the detector output is proportional to

$$|e^{i\Phi_1} + e^{i\Phi_2}|^2 = 2(1+\cos\Delta\Phi) \,, \tag{38}$$

where $\Phi_1(\Phi_2)$ refers to the phase accumulated along the first (second) beam and $\Delta\Phi = \Phi_1 - \Phi_2$. It is usually assumed that the two beams are coherently split at the source; therefore,

$$\Phi_1(\boldsymbol{x}_S, t_S) = \Phi_2(\boldsymbol{x}_S, t_S) \ . \tag{39}$$

We thus find

$$\hbar \ \Delta\Phi = -\int_{t_S}^{t_D} \Delta E \ dt + \oint \boldsymbol{p} \cdot d\boldsymbol{x} \ . \tag{40}$$

In stationary situations, it is possible to assume that $E_1 = E_2 = p_0^2/(2m)$, where (for $i = 1, 2$)

$$E_i = \frac{p_i^2}{2m} + \delta H_i \ . \tag{41}$$

Thus $\Delta E = 0$ and the calculation of the phase shift (40) can be simply performed if the perturbations δH_1 and δH_2 are small. It then follows from (41) that if $\delta\boldsymbol{p}$ is the perturbation in neutron momentum due to δH such that $\boldsymbol{p} - \delta\boldsymbol{p}$ is the "unperturbed" momentum with magnitude p_0, then

$$\boldsymbol{v} \cdot \delta\boldsymbol{p} = -\delta H \ , \tag{42}$$

where $\boldsymbol{v}$ is the neutron velocity. Hence, the *extra* phase shift due to the perturbation is given by

$$\Delta\Phi = \frac{1}{\hbar} \oint \delta\boldsymbol{p} \cdot d\boldsymbol{x} = \frac{1}{\hbar} \int_S^D (-\delta H_1 + \delta H_2) dt \ . \tag{43}$$

Consider, as an example, the Sagnac effect in the rotating frame, where $E = p^2/(2m) + \delta H$ with $\delta H = -\boldsymbol{\Omega} \cdot \boldsymbol{L}$. Thus (43) can be written as $\hbar \ \Delta\Phi = \oint \boldsymbol{\Omega} \cdot (m\boldsymbol{r} \times d\boldsymbol{r})$, since $\boldsymbol{L} = m\boldsymbol{r} \times \boldsymbol{v}$. In this way, one immediately recovers (25). The approach described here was originally employed for the calculation of the phase shift due to the spin-rotation coupling in a uniformly rotating system in [32].

In nonstationary situations, such as the proposed experiment using a rotating spin flipper, $\Delta E \neq 0$ and hence there is a beat phenomenon in addition to a phase shift. In fact, it follows from the analysis of that experiment [53] that $\Delta E = -\hbar\Omega$ for $t > t_{\text{out}}$, when the neutron exits the spin flippers. Hence $\Delta\Phi$ contains $\Omega(t_D - t_{\text{out}})$ in addition to a phase shift.

It is important to mention briefly the modification of spin-rotation coupling by the nonlocal theory of accelerated observers (Sect. 4). Equation (27) implies that E' can be positive, zero or negative. When $E' = 0$, the wave stands completely still with respect to the static observers in the rotating system. This is contrary to the basic postulate of the nonlocal theory; therefore, the only modification in (27) occurs for the $E' = 0$ case. This circumstance is discussed in detail in [20].

9 Translational Acceleration

Before treating quantum mechanics in translationally accelerated systems, it proves useful to digress here and discuss the transition from Lorentz invariance

to Galilean invariance in quantum mechanics. What is the transformation rule for a Schrödinger wave function under a Galilean boost ($t = t', \boldsymbol{x} = \boldsymbol{x}' + \boldsymbol{V}t$)? It follows from Lorentz invariance that for a spinless particle

$$\phi(x) = \phi'(x') \ , \tag{44}$$

where ϕ is a scalar wave function that satisfies the Klein-Gordon equation

$$\left(\Box + \frac{m^2c^2}{\hbar^2}\right)\phi(x) = 0 \ . \tag{45}$$

To obtain the Schrödinger equation from (45) in the nonrelativistic limit, we set

$$\phi(x) = \varphi(\boldsymbol{x}, t)e^{-i\frac{mc^2}{\hbar}t} \ . \tag{46}$$

Then, (45) reduces to

$$-\frac{\hbar^2}{2m}\nabla^2\varphi = i\hbar\frac{\partial\varphi}{\partial t} - \frac{\hbar^2}{2mc^2}\frac{\partial^2\varphi}{\partial t^2} \ . \tag{47}$$

Neglecting the term proportional to the second temporal derivative of φ in the nonrelativistic limit ($c \to \infty$), we recover the Schrödinger equation for the wave function φ.

Under a Lorentz boost, (44) and (46) imply that

$$\varphi(\boldsymbol{x}, t)e^{-i\frac{mc^2}{\hbar}t} = \varphi'(\boldsymbol{x}', t')e^{-i\frac{mc^2}{\hbar}t'} \ , \tag{48}$$

where

$$t = \gamma\left(t' + \frac{1}{c^2}\boldsymbol{V}\cdot\boldsymbol{x}'\right) \ . \tag{49}$$

It follows from

$$t - t' = \frac{1}{c^2}\left(\boldsymbol{V}\cdot\boldsymbol{x}' + \frac{1}{2}V^2t'\right) + O\left(\frac{1}{c^4}\right) \tag{50}$$

that in the nonrelativistic limit ($c \to \infty$),

$$\varphi(\boldsymbol{x}, t) = e^{i\frac{m}{\hbar}\left(\boldsymbol{V}\cdot\boldsymbol{x}' + \frac{1}{2}V^2t\right)}\varphi'(\boldsymbol{x}', t) \ . \tag{51}$$

This is the standard transformation formula for the Schrödinger wave function under a Galilean boost.

On the other hand, we expect from equations (2) and (44) that in the absence of spin, the wave function should turn out to be an invariant. Writing (48) in the form

$$\varphi(\boldsymbol{x}, t)e^{-i\frac{mc^2}{\hbar}t} = [\varphi'(\boldsymbol{x}', t')e^{i\frac{mc^2}{\hbar}(t-t')}]e^{-i\frac{mc^2}{\hbar}t} \ , \tag{52}$$

we note that the nonrelativisitic wave function may be assumed to be an invariant under a Galilean transformation

$$\psi(\boldsymbol{x}, t) = \psi'(\boldsymbol{x}', t) \ , \tag{53}$$

where

$$\psi(\boldsymbol{x},t)=\varphi(\boldsymbol{x},t),\quad \psi'(\boldsymbol{x}',t)=e^{i\frac{m}{\hbar}\left(\boldsymbol{V}\cdot\boldsymbol{x}'+\frac{1}{2}V^2t\right)}\varphi'(\boldsymbol{x}',t)\ . \tag{54}$$

That is, in this approach the phase factor in (51) that is due to the relativity of simultaneity belongs to the wave function itself.

The form invariance of the Schrödinger equation under Galilean transformations was used by Bargmann [55] to show that under the Galilei group, the wave function transforms as in (51). Bargmann used this result in a thought experiment involving the behavior of a wave function under the following four operations: a translation ($\boldsymbol{s}$) and then a boost ($\boldsymbol{V}$) followed by a translation ($-\boldsymbol{s}$) and finally a boost ($-\boldsymbol{V}$) to return to the original inertial system. It is straightforward to see from (51) that the original wave function $\varphi(\boldsymbol{x},t)$ is related to the final one $\varphi'(\boldsymbol{x},t)$ by

$$\varphi(\boldsymbol{x},t)=e^{-i\frac{m}{\hbar}\boldsymbol{s}\cdot\boldsymbol{V}}\varphi'(\boldsymbol{x},t)\ . \tag{55}$$

The phase factor in (55) leads to the *mass superselection rule*, namely, one cannot coherently superpose states of particles of different inertial masses [55, 56]. This rule guarantees strict conservation of mass in nonrelativistic quantum mechanics. The physical significance of this superselection rule has been critically discussed by Giulini [57] and more recently by Greenberger [58]. The main point here is that only Lorentz invariance is fundamental, since the nonrelativistic limit ($c\to\infty$) is never actually realized.

It should be clear from the preceding discussion that no mass superselection rule is encountered in the second approach based on the invariance of the wave function (53). It follows from the hypothesis of locality that the two distinct methods under discussion here carry over to the quantum mechanics of accelerated systems [59].

Let us therefore consider the transformation to an accelerated system

$$\boldsymbol{x}=\boldsymbol{x}'+\int_0^t\boldsymbol{V}(t')dt'\ , \tag{56}$$

where $\boldsymbol{a}=d\boldsymbol{V}/dt$ is the translational acceleration vector. Starting from the Schrödinger equation $H\psi=i\hbar\partial\psi/\partial t$ and assuming the invariance of the wave function, $\psi(\boldsymbol{x},t)=\psi'(\boldsymbol{x}',t)$, as in the second approach, we find that $\psi'(\boldsymbol{x}',t)=U\psi(\boldsymbol{x}',t)$, where

$$U=e^{\frac{i}{\hbar}\int_0^t\boldsymbol{V}(t')\cdot\boldsymbol{p}\,dt'}\ . \tag{57}$$

If follows that ψ' satisfies the Schrödinger equation $H'\psi'=i\hbar\partial\psi'/\partial t$ with the Hamiltonian

$$H'=UHU^{-1}-\boldsymbol{V}(t)\cdot\boldsymbol{p}\ , \tag{58}$$

where $\boldsymbol{p}$ is the invariant canonical momentum. Writing $H=p^2/(2m)+W$, where W is the invariant potential energy, we find

$$\left[\frac{1}{2m}(\boldsymbol{p}-m\boldsymbol{V})^2-\frac{1}{2}mV^2+W\right]\psi'=i\hbar\frac{\partial\psi'}{\partial t}\ . \tag{59}$$

Let

$$\psi'(\boldsymbol{x}',t) = e^{i\frac{m}{\hbar}\left[\boldsymbol{V}\cdot\boldsymbol{x}'+\frac{1}{2}\int_0^t V^2(t')dt'\right]}\varphi'(\boldsymbol{x}',t)\ , \tag{60}$$

then $\varphi'(\boldsymbol{x}',t)$ satisfies the Schrödinger equation

$$\left(-\frac{\hbar^2}{2m}\nabla'^2 + m\boldsymbol{a}\cdot\boldsymbol{x}' + W\right)\varphi' = i\hbar\frac{\partial\varphi'}{\partial t}\ , \tag{61}$$

where $\boldsymbol{\nabla}' = \boldsymbol{\nabla}$ follows from (56). It is important to recognize that $\varphi'(\boldsymbol{x}',t)$ is the wave function from the standpoint of the accelerated system according to the first (Bargmann) approach. Here the acceleration potential $m\boldsymbol{a}\cdot\boldsymbol{x}'$, where $-\boldsymbol{\nabla}'(m\boldsymbol{a}\cdot\boldsymbol{x}') = -m\boldsymbol{a}$ is the inertial force acting on the particle, corresponds to the inertial term that appears in (19). The existence of this inertial potential has been verified experimentally by Bonse and Wroblewski [60] using neutron interferometry. In connection with the problem of the wave function in the accelerated system – i.e. whether it is φ' or ψ' – a detailed examination of the experimental arrangement in [60] reveals that this experiment cannot distinguish between the two methods that differ by the phase factor given in (60). Specifically, the interferometer in [60] oscillated in the horizontal plane and the intensity of the outgoing beam was measured at the inversion points of the oscillation at which the magnitude of acceleration was maximum but $\boldsymbol{V} = 0$; therefore, the phase factor in question was essentially unity. To conclude our discussion, it is interesting to elucidate further the physical origin of this phase factor using classical mechanics [32].

Under the transformation (56), $\boldsymbol{v} = \boldsymbol{v}' + \boldsymbol{V}(t)$ and the Lagrangian of a classical particle $L = \frac{1}{2}mv^2 - W$, with $L(\boldsymbol{x},\boldsymbol{v}) = L'(\boldsymbol{x}',\boldsymbol{v}')$, becomes $L' = \frac{1}{2}m(\boldsymbol{v}'+\boldsymbol{V})^2 - W$ in the accelerated system. In classical mechanics, there are two natural and equivalent ways to deal with this Lagrangian. The first method consists of writing [31]

$$L' = \frac{1}{2}mv'^2 - m\boldsymbol{a}\cdot\boldsymbol{x}' - W + \frac{dF}{dt}\ , \tag{62}$$

where F is given, up to a constant, by

$$F = m\boldsymbol{V}(t)\cdot\boldsymbol{x}' + \frac{1}{2}m\int_0^t V^2(t')dt'\ . \tag{63}$$

The total temporal derivative in (62) does not affect the classical dynamics in accordance with the action principle and hence we confine our attention to $L'_1 = \frac{1}{2}mv'^2 - m\boldsymbol{a}\cdot\boldsymbol{x}' - W$. The momentum in this case is $\boldsymbol{p}' = m\boldsymbol{v}'$ and the Hamiltonian is thus given by

$$H'_1 = \frac{p'^2}{2m} + m\boldsymbol{a}\cdot\boldsymbol{x}' + W\ , \tag{64}$$

which corresponds to the Hamiltonian in the Schrödinger (61). The second method deals with L' without subtracting out dF/dt. In this case, the momentum is the invariant canonical momentum $\boldsymbol{p} = m(\boldsymbol{v}'+\boldsymbol{V})$ and the Hamiltonian is

$$H' = \frac{p^2}{2m} - \boldsymbol{p} \cdot \boldsymbol{V} + W \ , \tag{65}$$

which corresponds to (58) and the Hamiltonian in the Schrödinger (59).

In classical mechanics, the two methods represent the same dynamics. Quantum mechanically, however, there is a phase difference, which can be easily seen from the path integral approach. That is,

$$\psi'(\boldsymbol{x}', t) = \Sigma e^{\frac{i}{\hbar}\mathcal{S}'} \ , \tag{66}$$

where $\mathcal{S}'$ is the classical action,

$$\mathcal{S}' = \int L'(\boldsymbol{x}', \boldsymbol{v}')dt \ . \tag{67}$$

It follows from (62) that

$$\mathcal{S}' = \mathcal{S}'_1 + F \ , \tag{68}$$

where $\mathcal{S}'_1$ is the action corresponding to L'_1. Using (68) and the fact that

$$\varphi'(\boldsymbol{x}', t) = \Sigma e^{\frac{i}{\hbar}\mathcal{S}'_1} \ , \tag{69}$$

we find

$$\psi'(\boldsymbol{x}', t) = e^{\frac{i}{\hbar}F}\varphi'(\boldsymbol{x}', t) \ , \tag{70}$$

in agreement with (60).

It would be interesting to devise an experiment of the Bonse-Wroblewski [60] type that could distinguish between the two methods and hence remove the phase ambiguity in the treatment of translationally accelerated systems.

10 Discussion

The main observational consequences of Dirac's equation in noninertial frames of reference are related to the Sagnac effect, the spin-rotation coupling and the Bonse-Wroblewski effect. These inertial effects can be further elucidated by interferometry experiments involving matter waves. In particular, a neutron interferometry experiment has been proposed for the direct measurement of inertial effect of intrinsic spin. Moreover, neutron interferometry experiments involving translationally accelerated interferometers may help resolve the phase ambiguity in the description of the wave function from the standpoint of a translationally accelerated system.

References

1. P.A.M. Dirac, Proc. Roy. Soc. (London) A**117**, 610 (1928); **118**, 351 (1928).
2. V.A. Fock and D.D. Ivanenko, C.R. Acad. Sci. **188**, 1470 (1929); Z. Phys. **54**, 798 (1929).

3. H. Tetrode, Z. Phys. **50**, 336 (1928); V. Bargmann, Sitzber., Preuss. Akad. Wiss. Phys.-Math. Kl. 346 (1932); E. Schrödinger, Sitzber., Preuss. Akad. Wiss. Phys.-Math. Kl. 105 (1932); L. Infeld and B.L. van der Waerden, Sitzber., Preuss. Akad. Wiss. Phys.-Math. Kl. 380 (1933); 474 (1933); D.R. Brill and J.A. Wheeler, Rev. Mod. Phys. **29**, 465 (1957); **33**, 623E (1961).
4. H.A. Lorentz, *The Theory of Electrons* (Dover, New York, 1952).
5. A. Einstein, *The Meaning of Relativity* (Princeton University Press, Princeton, 1950).
6. B. Mashhoon, Phys. Lett. A **143**, 176 (1990); **145**, 147 (1990).
7. H. Weyl, *Space-Time-Matter* (Dover, New York, 1952), pp. 176–177.
8. H. Minkowski, in *The Principle of Relativity*, by H.A. Lorentz, A. Einstein, H. Minkowski and H. Weyl (Dover, New York, 1952).
9. B. Mashhoon, in *Relativity in Rotating Frames*, edited by G. Rizzi and M.L. Ruggiero (Kluwer, Dordrecht, 2004), pp. 43–55.
10. B. Mashhoon, Phys. Rev. A **47**, 4498 (1993); in *Cosmology and Gravitation*, edited by M. Novello (Editions Frontières, Gif-sur-Yvette, 1994), pp. 245-295; U. Muench, F.W. Hehl and B. Mashhoon, Phys. Lett. A **271**, 8 (2000).
11. N. Bohr and L. Rosenfeld, K. Dan. Vindensk. Selsk. Mat. Fys. Medd. **12**, No. 8, (1993); translated in *Quantum Theory and Measurement*, edited by J.A. Wheeler and W.H. Zurek (Princeton University Press, Princeton, 1983); Phys. Rev. **78**, 794 (1950).
12. V. Volterra, *Theory of Functionals and of Integral and Integro-Differential Equations* (Dover, New York, 1959).
13. F.G. Tricomi, *Integral Equations* (Interscience, New York, 1957).
14. C. Chicone and B. Mashhoon, Ann. Phys. (Leipzig) **11**, 309 (2002).
15. C. Chicone and B. Mashhoon, Phys. Lett. A **298**, 229 (2002).
16. F.W. Hehl and Y.N. Obukhov, *Foundations of Classical Electrodynamics* (Birkhäuser, Boston, 2003).
17. B. Mashhoon, Ann. Phys. (Leipzig) **12**, 586 (2005); Int. J. Mod. Phys. D **14**, 171 (2005).
18. R. Sauerbrey, Phys. Plasmas **3**, 4712 (1996); G. Schäfer and R. Sauerbrey, astro-ph/9805106.
19. B. Mashhoon, Phys. Rev. A **70**, 062103 (2004).
20. B. Mashhoon, Phys. Rev. A **72**, 052105 (2005).
21. C.G. de Oliveira and J. Tiomno, Nuovo Cimento **24**, 672 (1962); J. Audretsch and G. Schäfer, Gen. Relativ. Gravit. **9**, 243 (1978); E. Fischbach, B.S. Freeman and W.-K. Cheng, Phys. Rev. D **23**, 2157 (1981); F.W. Hehl and W.-T. Ni, Phys. Rev. D **42**, 2045 (1990); Y.Q. Cai and G. Papini, Phys. Rev. Lett. **66**, 1259 (1991); **68**, 3811 (1992); J.C. Huang, Ann. Phys. (Leipzig) **3**, 53 (1994); K. Konno and M. Kasai, Prog. Theor. Phys. **100**, 1145 (1998); K. Varjú and L.H. Ryder, Phys. Lett. A **250**, 263 (1998); L. Ryder, J. Phys. A **31**, 2465 (1998); K. Varjú and L.H. Ryder, Phys. Rev. D **62**, 024016 (2000).
22. N.V. Mitskievich, *Physical Fields in General Relativity Theory*, in Russian (Nauka, Moscow, 1969); E. Schmutzer, Ann. Phys. (Leipzig) **29**, 75 (1973); B.M. Barker and R.F. O'Connell, Phys. Rev. D **12**, 329 (1975); T.C. Chapman and D.J. Leiter, Am. J. Phys. **44**, 858 (1976); E. Schmutzer and J. Plebański, Fortschr. Phys. **25**, 37 (1977).
23. J. Audretsch and C. Lämmerzahl, Appl. Phys. B **54**, 351 (1992); J. Audretsch, F.W. Hehl and C. Lämmerzahl, Lect. Notes Phys. **410**, 368 (1992); C. Lämmerzahl, Gen. Rel. Grav. **28**, 1043 (1996); Class. Quantum Grav. **15**, 13 (1998).

24. J. Anandan, Phys. Rev. Lett. **68**, 3809 (1992); J. Anandan and J. Suzuki, in *Relativity in Rotating Frames*, edited by G. Rizzi and M.L. Ruggiero (Kluwer, Dordrecht, 2004), pp. 361–370; Y.Q. Cai, D.G. Lloyd and G. Papini, Phys. Lett. A **178**, 225 (1993); D. Singh and G. Papini, Nuovo Cimento B **115**, 223 (2000); G. Papini, in *Relativity in Rotating Frames*, edited by G. Rizzi and M.L. Ruggiero (Kluwer, Dordrecht, 2004), pp. 335-359; Z. Lalak, S. Pokorski and J. Wess, Phys. Lett. B **355**, 453 (1995); I. Damiao Soares and J. Tiomno, Phys. Rev. D **54**, 2808 (1996); C.J. Bordé, J.-C. Houard and A. Karasiewicz, Lect. Notes Phys. **562**, 403 (2001); C. Kiefer and C. Weber, Ann. Phys. (Leipzig) **14**, 253 (2005).
25. F.W. Hehl and W.-T. Ni, Phys. Rev. D **42**, 2045 (1990).
26. L.L. Foldy and S.A. Wouthuysen, Phys. Rev. **78**, 29 (1950).
27. Y.N. Obukhov, Phys. Rev. Lett. **86**, 192 (2001); Fortschr. Phys. **50**, 711 (2002); N. Nicolaevici, Phys. Rev. Lett. **89**, 068902 (2002); Y.N. Obukhov, Phys. Rev. Lett. **89**, 068903 (2002).
28. M.V. Romalis, W.C. Griffith, J.P. Jacobs and E.N. Fortson, Phys. Rev. Lett. **86**, 2505 (2001).
29. D. Bini, C. Cherubini and B. Mashhoon, Class. Quantum Grav. **21**, 3893 (2004).
30. A.J. Silenko and O.V. Teryaev, Phys. Rev. D **71**, 064016 (2005).
31. L.D. Landau and E.M. Lifshitz, *Mechanics* (Pergamon, Oxford, 1966).
32. B. Mashhoon, Phys. Rev. Lett. **61**, 2639 (1988); **68**, 3812 (1992).
33. J.J. Sakurai, Phys. Rev. D **21**, 2993 (1980).
34. J.E. Zimmerman and J.E. Mercereau, Phys. Rev. Lett. **14**, 887 (1965).
35. S.A. Werner, J.-L. Staudenmann and R. Colella, Phys. Rev. Lett. **42**, 1103 (1979).
36. D.K. Atwood, M.A. Horne, C.G. Shull and J. Arthur, Phys. Rev. Lett. **52**, 1673 (1984).
37. F. Riehle, T. Kisters, A. Witte, J. Helmcke and C.J. Bordé, Phys. Rev. Lett. **67**, 177 (1991).
38. A. Lenef et al., Phys. Rev. Lett. **78**, 760 (1997); T.L. Gustavson, P. Bouyer and M.A. Kasevich, Phys. Rev. Lett. **78**, 2046 (1997); T.L. Gustavson, A. Landragin and M.A. Kasevich, Class. Quantum Grav. **17**, 2385 (2000).
39. F. Hasselbach and M. Nicklaus, Phys. Rev. A **48**, 143 (1993).
40. E.J. Post, Rev. Mod. Phys. **39**, 475 (1967); G.E. Stedman, Rep. Prog. Phys. **60**, 615 (1997).
41. L. Ryder, J. Phys. A **31**, 2465 (1998); L.H. Ryder and B. Mashhoon, in *Proc. Ninth Marcel Grossmann Meeting*, edited by V.G. Gurzadyan, R.T. Jantzen and R. Ruffini (World Scientific, Singapore, 2002), pp. 486–497.
42. B. Mashhoon, Found. Phys. **16** (Wheeler Festschrift), 619 (1986).
43. B. Mashhoon, Phys. Lett. A **173**, 347 (1993); J.C. Hauck and B. Mashhoon, Ann. Phys. (Leipzig) **12**, 275 (2003).
44. N. Ashby, Living Rev. Relativity **6**, 1 (2003).
45. B. Mashhoon, Phys. Lett. A **306**, 66 (2002).
46. J.D. Anderson and B. Mashhoon, Phys. Lett. A **315**, 199 (2003).
47. P.J. Allen, Am. J. Phys. **34**, 1185 (1966).
48. B. Mashhoon, Phys. Lett. A **198**, 9 (1995).
49. B.J. Venema et al., Phys. Rev. Lett. **68**, 135 (1992); D.J. Wineland et al., Phys. Rev. Lett. **67**, 1735 (1991).
50. B. Mashhoon, Gen. Rel. Grav. **31**, 681 (1999); Class. Quantum Grav. **17**, 2399 (2000).
51. G. Papini, Phys. Rev. D **65**, 077901 (2002); G. Papini and G. Lambiase, Phys. Lett. A **294**, 175 (2002); G. Lambiase and G. Papini, Phys. Rev. D **70**, 097901 (2004); D. Singh, N. Mobed and G. Papini, J. Phys. A **37**, 8329 (2004).

52. H. Rauch and S.A. Werner, *Neutron Interferometry* (Oxford University Press, New York, 2000).
53. B. Mashhoon, R. Neutze, M. Hannam and G.E. Stedman, Phys. Lett. A **249**, 161 (1998).
54. G. Badurek, H. Rauch and D. Tuppinger, Phys. Rev. A **34**, 2600 (1986).
55. V. Bargmann, Ann. Math. **59**, 1 (1954).
56. F.A. Kaempffer, *Concepts in Quantum Mechanics* (Academic Press, New York, 1965).
57. D. Giulini, Ann. Phys. (NY) **249**, 222 (1996).
58. D.M. Greenberger, Phys. Rev. Lett. **87**, 100405 (2001).
59. S. Takagi, Prog. Theor. Phys. **86**, 783 (1991); W.H. Klink, Ann. Phys. (NY) **260**, 27 (1997); I. Bialynicki-Birula and Z. Bialynicka-Birula, Phys. Rev. Lett. **78**, 2539 (1997).
60. U. Bonse and T. Wroblewski, Phys. Rev. Lett. **51**, 1401 (1983); Phys. Rev. D **30**, 1214 (1984).

Vacuum Fluctuations, Geometric Modular Action and Relativistic Quantum Information Theory

R. Verch

Institut für Theoretische Physik, Universität Leipzig Postfach 10 09 20, 04009 Leipzig, Germany
verch@itp.uni-leipzig.de

Abstract. A summary of some lines of ideas leading to model-independent frameworks of relativistic quantum field theory is given. It is followed by a discussion of the Reeh-Schlieder theorem and geometric modular action of Tomita-Takesaki modular objects associated with the quantum field vacuum state and certain algebras of observables. The distillability concept, which is significant in specifying useful entanglement in quantum information theory, is discussed within the setting of general relativistic quantum field theory.

1 Introduction

About 100 years ago, new insights into the physical world were gained which at that time had a new quality to them. The new feature was that certain phenomena could successfully be described by means of concepts which have little in common with the behavior of physical objects familiar from everyday experience. The first of these insights we are referring to was Planck's quantum hypothesis in his account of black-body radiation. The second was Einstein's theory of special relativity. (See, e.g., [45] for a historical presentation of these developments.)

It took a while – more or less, two decades – until quantum theory reached the form of (non-relativistic) quantum mechanics which is nowadays taught in courses at universities. A further step was the combination and unification of the principles of quantum mechanics and special relativity. The endeavors to accomplish this step took still longer – and, rigorously speaking, they haven't come to an end even today. And the synthesis of quantum mechanics and general relativity into some form of a quantum theory of gravity lies still well ahead of us.

The theory unifying the principles of quantum mechanics and special relativity has come to be called *relativistic quantum field theory*, or QFT, for short. To delineate the basic characteristics of QFT, let us recall first the basic features of

R. Verch: *Vacuum Fluctuations, Geometric Modular Action and Relativistic Quantum Information Theory*, Lect. Notes Phys. **702**, 133–162 (2006)
DOI 10.1007/3-540-34523-X_6 © Springer-Verlag Berlin Heidelberg 2006

Quantum mechanics, which provides a conceptual foundation for describing physical processes at small scales (in space and time), and is therefore relevant in the microscopic domain and accounts for the stability of atoms and molecules. Moreover, its (experimentally testable) predictions are of statistical nature, with the characteristic feature of uncertainty relations.

Special relativity, on the other hand, can be viewed as providing a conceptual foundation for the description of space and time, relevant in particular in the context of processes involving very high energies and momenta. Among its principal features are the absence of preferred inertial frames (observers), i.e. Poincaré-covariance, the speed of light as maximal velocity of signal propagation, and matter (mass)-energy equivalence.

The fundamental aspects of both quantum mechanics and special relativity find a unification in the form of

Quantum field theory, which consequently provides a theoretical framework for the description of processes with very high energy/momentum exchange at very small time/length scales; it is therefore relevant in the submicroscopic domain and accounts for the properties and the stability of elementary particles, predicts annihilation and creation of particles, new types of charges, anti–charges, PCT and spin–statistics theorems, fluctuations and long–range correlations.

While this is not the place to give a review of the historical development of QFT and its interplay with the development of elementary particle physics, involving also new concepts such as renormalization, internal group symmetries, gauge theory, spontaneous symmetry breaking, Higgs mechanism etc., there are some comments to be made at this point about the various sub-branches of QFT and its status as a physical theory, as well as its status as concerns mathematical consistency of the framework.

Let us begin by mentioning the by far largest branch of QFT, which we refer to as *perturbative QFT*. The idea here is to look at concrete *quantum field models*, mostly in the form of a Lagrangean for an – initially – classical field theory model involving certain types of matter and gauge fields. Typically, the fields interact in some way and this leads to the occurrence of multilinear (polynomial) expressions of the fields in the field equations. One would then like to have "quantized" solutions of the field equations. It is not a priori clear what this means, but the pragmatic way to proceed is as follows. One starts with the interaction-free part of the field equation (neglecting the multilinear, interacting parts of the field equations) and constructs "quantized" solutions for that in the form of "free" quantum fields – where it is in most of the relevant cases known what this means. Then one regards the interacting expressions of the (now quantized, free) fields as a perturbation of the free dynamics, and tries to construct solutions to the full dynamics by means of a perturbation series in the parameter specifying the strength of the interaction (the coupling parameter). At this point there arises the difficulty that the various multilinear expressions in the fields appearing in the perturbation series are not well defined at the level

of (free) quantized fields, and that they need to be "renormalized". If this is possible systematically to all polynomial orders upon introducing only finitely many parameters (to be determined experimentally), one calls the quantum field model under consideration *(perturbatively) renormalizable*. Once the renormalization parameters are determined experimentally, predictions of the quantum field model can be compared with experimental data e.g. obtained in scattering experiments with elementary particles – up to a given order in the coupling parameter of the perturbation series.

The successes of perturbative quantum field theory in comparison with experiment are truly impressive. The numerical agreement of theoretical predictions and experimental data is in many cases of the in the range of 8 significant figures or better, and also properties of particles whose existence was predicted by QFT prior to observation, like in the case of the $W^{\pm}$ and Z^0 bosons in the electroweak interactions, are in excellent agreement with experimental findings. (See [34, 68] for the various aspects of perturbative QFT.)

However, from a more fundamental point of view, perturbative QFT is not fully satisfactory. The perturbation series by which one attempts to approximate the full interacting quantum field dynamics won't converge, and then it is unclear if there is a solution to the quantized field equations at all. This provokes the question at which order in the coupling parameter the perturbation series ought to be truncated to yield acceptable agreement with experiment, and this question remains so far unanswered within perturbative quantum field theory. Moreover, the number of renormalization parameters which have to be determined by experiment and are not derivable within perturbation theory are quite large for physically realistic quantum field models (of the order of about 20 in the case of the standard model), and this is regarded as a considerable drawback as concerns the predictive power of perturbative QFT.

Hence, there clearly is room for approaches to QFT (and elementary particle physics) other than by perturbative QFT. Let me point out three basic branches. One idea is that theories such as the standard model are simply not rich enough and/or do not include all interactions (such as gravity), and that a richer theory should be considered in the first place (first at the level of a "classical field theory" then quantized, maybe at the level of perturbative QFT), with the hope that the richer symmetry structure constrains the amount of free parameters considerably. *Grand unified theories*, and *string theory*, can be seen in this light.

The next branch is *constructive quantum field theory*, where one attempts to construct solutions to the quantized, interacting field equations mathematically rigorously. This branch of QFT is much smaller than those mentioned previously, but has had quite impressive successes which are partly documented in [27, 53]. The mathematical difficulties one is faced with in constructive QFT are immense, not least by the circumstance that it is often not entirely clear what is actually meant by a solution to a quantized, interacting field equation (we will soon come back to this point). Nevertheless, interacting quantum field models have been rigorously constructed in spacetime dimensions 2 and 3. The case of a rigorous solution to quantized field equations for models regarded as physically relevant

remains open in 4 spacetime dimensions and is still an area of active research. The Clay Institute of Mathematics awards a million dollars for the solution of this problem. (There is also a branch of QFT which is known as *lattice gauge theory*, and which can be placed somewhere between perturbative QFT and constructive QFT. The interested reader is referred to [44] for more information about it.)

Finally, there is yet another branch of QFT, commonly called *axiomatic quantum field theory*, although this labeling is to some degree misleading. The basic idea is that one wishes to formulate and analyze the properties which are thought to be common to all physically realistic quantum field models. This is on one hand indispensable to make the problem of rigorous construction of interacting quantum field models a mathematically well-defined problem, on the other hand it is also difficult in the absence of rigorously constructed interacting quantum field models in 4 spacetime dimensions as a guidance. To begin with, the task is to find a mathematical structure which encodes the basic principles of quantum mechanics and special relativity, and which subsumes the known rigorously constructed quantum field models where these principles are implemented (e.g. for free quantum fields, or for interacting quantum fields in lower spacetime dimension). This task was taken up initially by Wightman and others (see [8, 37, 60]) from a point of view involving mainly distribution theory, and by Haag and Kastler [28, 29] using the mathematical theory of operator algebras. Seen from a mathematician's perspective, the latter approach turned out to be more fruitful. In fact, there are many rigorous and deep results about the mathematical structure of (model-independent) quantum field theory in the operator algebraic framework. The reader might like to consult [1, 3, 28] for a comprehensive review.

The present contribution is, in fact, placed within the framework of axiomatic QFT. In the next section, we will sketch how one can combine the principles of quantum mechanics and of special relativity in a mathematical structure which more or less is "common to all quantum field models". Then we will present the "Reeh–Schlieder–theorem" and discuss some of aspects of it. The Reeh–Schlieder–theorem is a strong mathematical statement about the ubiquity and complexity of vacuum fluctuations in quantum field theory, regardless of the particular quantum field theoretical model considered: It is a consequence of first principles such as locality (causal propagation), stability of the vacuum, and covariance. Then we will discuss a mathematical structure arising in connection with the Reeh-Schlieder-theorem: Geometric modular action. While discovered already in 1975 by Bisognano and Wichmann [7], this mathematical structure has in the recent years given rise to many new insights into quantum field field theory which we will briefly discuss. In a sense, it unifies the mathematical domains of quantum mechanics – operator algebras – and of special relativity – affine geometry – completely. Moreover, it opens very interesting new perspectives.

We will then proceed to another topic where the Reeh-Schlieder-theorem plays again a prominent role: In discussing aspects of entanglement in the framework of relativistic QFT. This part of the present contribution is essentially a

summary of parts of a recent joint work with R. Werner [67]. We will present a variant of the distillability concept of bipartite systems in quantum field theory. Furthermore, we will quote our result stating that the vacuum state (as well as any relativistic thermal equilibrium state) is distillable over arbitrary spacelike distances.

Taking up a line of thought mentioned at the very beginning of this introduction, we should like to point out that also in the realm of phenomena described by quantum field theory one encounters theoretical propositions which at first sight appear implausible because of their highly counterintuitive character. The Reeh–Schlieder–theorem serves as an example, as well as distillability of the vacuum state. However, careful statement of the concepts and careful analysis of their consequences, together with proper use of adequate mathematical methods, will bring us closer to an understanding of these novel situations and, ultimately, their experimental testing. Thus, we will need to collect also some mathematical concepts and results which are not necessarily in every theoretical physicist's toolbox. Nevertheless, we have tried to keep the amount of formalities at a minimum and to make this contribution as self-contained as possible, hoping that everyone familiar with quantum mechanics, special relativity and the rudiments of quantum field theory will be able to follow this contribution without undue strain.

2 From Quantum Mechanics and Special Relativity to Quantum Field Theory

Let us once more recall the basic features of quantum mechanics, this time at a more formal level. The theory of quantum mechanics says that a quantum mechanical system is described by:

- $\mathcal{H}$: a Hilbert space
- $\mathcal{R} \subset B(\mathcal{H})$: a $*$-algebra of operators, where:
  - $A = A^* \in \mathcal{R}$ is interpreted as an *observable*
  - For $\psi \in \mathcal{H}$ with $||\psi|| = 1$, the quantity

$$\langle A \rangle_\psi = \langle \psi, A\psi \rangle$$

    is interpreted as the *expectation value* of the observable A in the *state* given by ψ. More generally: For $\rho =$ trace-class operator on $\mathcal{H}$ with $\rho \geq 0$, $\mathrm{trace}(\rho) = 1$, we interpret $\langle A \rangle_\rho = \mathrm{trace}(\rho\, A)$ as expectation value of A in the state given by ρ.

We need to explain some notation and terminology appearing here. First note that by Hilbert space we mean a complex-linear Hilbert space. The scalar product of two vectors $\psi, \phi \in \mathcal{H}$ is denoted $\langle \psi, \phi \rangle$, and $||\psi||^2 = \langle \psi, \psi \rangle$. By $B(\mathcal{H})$ we denote the set of all bounded linear operators $A : \mathcal{H} \to \mathcal{H}$. A subset $\mathcal{R}$ of $B(\mathcal{H})$ (which may, but need not, coincide with $B(\mathcal{H})$) is a $*$-algebra if, given A and B in $\mathcal{R}$ and $\lambda, \mu \in \mathbb{C}$, the operators $\lambda A + \mu B$, AB and A^* are again

contained in $\mathcal{R}$, where A^* is the adjoint operator. Hence, a quantum mechanical system is described by specifying its state Hilbert space $\mathcal{H}$ and its algebra of observables $\mathcal{R}$.

There are a few remarks to be made:

(2.1) One might take the point of view that the description of a quantum mechanical system requires also the specification of dynamics, e.g. in the form of a Hamiltonian operator H acting in $\mathcal{H}$. Furthermore, one may also require that the quantum system admits states of lowest energy for H ("ground states") [or that the spectrum of H is bounded below], or thermal equilibrium states, since the sudden decay of matter which would otherwise occur (for quantum systems not having these properties) is not observed in real systems. We shall ignore aspects of dynamics for the moment, but will come back to this point later in the discussion of quantum field theory.

(2.2) It is tacitly assumed that $\mathcal{R}$ is non-abelian, i.e. that $AB \neq BA$ holds for some A and B in $\mathcal{R}$, as otherwise there are no uncertainty relations which are characteristic of quantum theory.

(2.3) One may wonder if the setting presented here is general enough since $\mathcal{R}$ contains only bound operators, while in quantum mechanics of single particles observables like position or momentum are represented by unbounded operators as a consequence of the canonical commutation relations. Employing the functional calculus, however, one may pass e.g. from the unbounded operator P representing the observable "momentum" to the bounded operator $f(P)$, which is bounded when f is a bounded real function on $\mathcal{R}$, and which represents the observable "f(momentum)". This shows that it is in general no loss of physical information to work only with bound operators as observables; moreover, unbounded operators can be regarded as suitable limits of sequences of bounded operators. Working with bounded operators has considerable advantages as far as the mathematical analysis is concerned, since subtle domain problems that plague the rigorous manipulation of unbounded operators are avoided.

(2.4) One may also wonder why we have not simply taken $\mathcal{R} = B(\mathcal{H})$, the standard case in quantum mechanics of a single particle. The reason is that we would like to allow greater flexibility, making it possible to consider also subsystems of a larger, ambient system. An example, occurring often in quantum information theory, is the case $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$ with $\mathcal{R} = B(\mathcal{H}_1) \otimes \mathbf{1}$ modeling a subsystem of the full system whose algebra of observables is given by $B(\mathcal{H}_1 \otimes \mathcal{H}_2) \supset B(\mathcal{H}_1) \otimes B(\mathcal{H}_2)$. We will encounter a similar situation later. In discussions of model-independent properties of quantum field theories, $\mathcal{R}$ often means the algebra of observables measurable – and in this sense, localized – in a proper subregion of Minkowski spacetime, as we will discuss below.

Having thus collected the basics of the formal framework of quantum mechanics, we turn now to special relativity. We will be very brief in recalling its

basic formal ingredients. The theory of special relativity states that all physical events can be collected in a catalogue which has the structure of a 4–dimensional affine space M, where each point in M represents a (possible) event. There is a metric η of Lorentzian signature on M; that is, one can choose identifications of M with $\mathbb{R}^4$ in such a way that, with respect to the standard coordinates of $\mathbb{R}^4$, η is represented by the diagonal matrix $\mathrm{diag}(1,-1,-1,-1)$. The choice of such an identification is also referred to as fixing of an inertial frame. With respect to a fixing of an inertial frame (inducing an orientation an a time–orientation), one can introduce the proper orthochronous Poincaré group $\mathfrak{P}_+^\uparrow$, which is the unit connected component of the full Poincaré group $\mathfrak{P}$, defined as the group of all invertible affine transformations of M leaving η invariant. We assume from now on that an inertial frame has been fixed. Any $L \in \mathfrak{P}$ (or $\mathfrak{P}_+^\uparrow$) decomposes as a semidirect product of $\Lambda \in \mathfrak{L}$ (or $\mathfrak{L}_+^\uparrow$), the Lorentz group (or its unit connected component) and $a \in \mathfrak{T} \equiv \mathbb{R}^4$, the group of translations, according to

$$Lx = (\Lambda, a)x = \Lambda x + a\,, \quad x \in M \equiv \mathbb{R}^4\,.$$

The reader is referred to the contribution by D. Giulini in this volume for a full discussion of special relativity, Minkowski spacetime and the Poincaré group (alternatively, see e.g. [59]).

The theory of special relativity states that the description of a physical system is equivalent for all inertial observers, i.e. in arbitrary inertial frames. Put differently, the description of physical processes should be covariant with respect to proper, orthochronous Poincaré transformations. More formally, this means:

> Suppose a quantum system is modeled by $(\mathcal{R}, \mathcal{H})$. Let ρ be a density matrix and A an observable with respect to a given inertial frame. If $L \in \mathfrak{P}_+^\uparrow$, then there corresponds, with respect to the L-transformed inertial frame, a density matrix ρ_L and observable A_L to ρ and A, respectively, such that
>
> $$\langle A_L \rangle_{\rho_L} = \langle A \rangle_\rho\,. \tag{1}$$

One can add some mathematical precision, requiring that the maps taking A to A_L and ρ to ρ_L are one–to–one and onto, i.e. bijective. Following Wigner, one may think of elementary systems where $\mathcal{R} = B(\mathcal{H})$, and then one can conclude:

> There is a unitary representation
>
> $$\widetilde{\mathfrak{P}}_+^\uparrow \ni \widetilde{L} \mapsto \widetilde{U}(\widetilde{L})$$
>
> of the universal covering group of $\mathfrak{P}_+^\uparrow$ on $\mathcal{H}$, such that
>
> $$A_L = \widetilde{U}(\widetilde{L}) A \widetilde{U}(\widetilde{L})^*\,, \quad \rho_L = \widetilde{U}(\widetilde{L}) \rho \widetilde{U}(\widetilde{L})^*\,,$$
>
> where $\widetilde{\mathfrak{P}}_+^\uparrow \ni \widetilde{L} \mapsto L \in \mathfrak{P}_+^\uparrow$ is the canonical projection. Moreover, if suitable assumptions about the continuity of the maps $A \mapsto A_L$, $\rho \mapsto \rho_L$ are made – and we tacitly make this assumption – then one can conclude that the unitaries $\widetilde{U}(\widetilde{L})$ depend continuously on $\widetilde{L}$.

This result is known as the Wigner–Bargmann–theorem, which actually holds under somewhat weaker assumptions than expressed in (1); it is sufficient to consider as observables 1–dimensional projections $A = |\psi\rangle\langle\psi|$ and likewise, 1–dimensional projections $\rho = |\phi\rangle\langle\phi|$ as density matrices, and to replace (1) by the weaker requirement

$$|\langle A_L\rangle_{\rho_L}| = |\langle A\rangle_\rho| \,.$$

We refer to the original articles by Wigner [70] and Bargmann [2] and to [28,55, 60] for considerable further discussion.

The Wigner–Bargmann–theorem states that, in the case of an (elementary) quantum system compatible with the covariance principle of special relativity, the state Hilbert space $\mathcal{H}$ carries a unitary representation of $\widetilde{\mathfrak{P}}^\uparrow_+$, the universal covering group of the proper orthochronous Poincaré group, implementing the change of inertial frames. The appearance of a unitary representation of the universal covering group instead of the proper orthochronous Poincaré group itself is due to the fact that (1) fixes only a unitary representation of $\mathfrak{P}^\uparrow_+$ up to a phase, but this can be lifted to a proper unitary representation of $\widetilde{\mathfrak{P}}^\uparrow_+$. The significance of this was clarified by Wigner's analysis of the irreducible unitary representations of $\widetilde{\mathfrak{P}}^\uparrow_+$ having positive energy, thereby making the term "elementary system" precise. The Hilbert spaces supporting these irreducible unitary representations ("one–particle spaces") correspond to spaces of solutions of linear wave equations, like the Klein–Gordon, Dirac or Maxwell equations in the simplest cases. The mass and the spin (or helicity) of these wave equations is a distinguishing label for the irreducible unitary representations of $\widetilde{\mathfrak{P}}^\uparrow_+$.

Wigner's analysis reveals some structural elements of quantum mechanical systems compatible with the principles of special relativity, but not all, in particular the aspect of a "quantized field" hasn't appeared yet. To see how this aspect comes into play, one usually takes a complementary route: Consider a typical $\mathfrak{P}^\uparrow_+$–covariant classical system; i.e. a classical field subject to a linear wave-equation. The electromagnetic field provides the prime and archetypical example, but let us consider here a much simpler example, the scalar Klein-Gordon field $\varphi(x)$, $x \in M \equiv \mathbb{R}^4$, obeying the following equation of motion:

$$\left(\eta^{\mu\nu}\frac{\partial}{\partial x^\mu}\frac{\partial}{\partial x^\nu} + m^2\right)\varphi(x) = 0$$

where $m \geq 0$ is a constant. Such a classical field can be viewed as a Hamiltonian system with infinitely many degrees of freedom, and one may therefore try and quantize it by regarding it as a "limit" of a Hamiltonian system with N degrees of freedom as $N \to \infty$, and taking as its quantized version the "limit" of the quantized systems with N degrees of freedom as $N \to \infty$. In the case of the Klein-Gordon field, the classical field $\varphi(x^0, \boldsymbol{x})$, $x = (x^0, \boldsymbol{x}) \in \mathbb{R}\times\mathbb{R}^3$, at time-coordinate x^0 (with respect to an arbitrary but fixed inertial frame) can be approximated by a discrete lattice of coupled harmonic oscillators with canonical coordinates $q_{\lambda\mu\nu}(x^0)$ at the lattice site

$$\boldsymbol{x}(\lambda,\mu,\nu) = a \begin{bmatrix} \lambda \\ \mu \\ \nu \end{bmatrix} \in \mathbb{R}^3\,, \quad \lambda,\mu,\nu \in \mathbb{Z}\,, \quad |\lambda|,|\mu|,|\nu| \leq \frac{1}{a^2}\,,$$

where $a > 0$ is the lattice spacing. To the discrete lattice system one can associate the quantum system of coupled harmonic oscillators (at lattice spacing a, there are $N \sim 1/a^6$ of them), where the canonical classical coordinates $q_{\lambda\mu\nu}(x^0)$ and conjugate momenta $p_{\lambda\mu\nu}(x^0)$ become operators $Q_{\lambda\mu\nu}(x^0)$ and $P_{\lambda\mu\nu}(x^0)$ obeying the canonical commutation relations. In the limit as $a \to 0$ and $N \to \infty$, one obtains for each $f, h \in C_0^\infty(\mathbb{R}^3)$ the field operators

$$\begin{aligned} \Phi(x^0,f) &= \lim_{a\to 0,\, N\to\infty} \sum_{\lambda,\mu,\nu} Q_{\lambda\mu\nu}(x^0) f(\boldsymbol{x}(\lambda,\mu,\nu)) a^3\,, \\ \Pi(x^0,f) &= \lim_{a\to 0,\, N\to\infty} \sum_{\lambda,\mu,\nu} P_{\lambda\mu\nu}(x^0) f(\boldsymbol{x}(\lambda,\mu,\nu))\,. \end{aligned}$$

For a detailed discussion of this construction, cf. [33]. To summarize, we find the following formal correspondences (where we use the shorthand j or ℓ for the index triple $\lambda\mu\nu$, and occasionally drop the time–argument x^0):

Classical Mechanics	**Quantum Mechanics**
Phase-space fncts $q_1, \ldots, q_k, p_1, \ldots, q_k$	Operators $Q_1, \ldots, Q_k, P_1, \ldots, P_k$
Poisson brackets: $\{q_j, p_\ell\} = \delta_{j\ell}$	Commutators: $[Q_j, P_\ell] = i\hbar\delta_{j\ell}$
Classical Field Theory	**Quantum Field Theory**
field can. conj. momentum $\varphi(x^0,\boldsymbol{x})$ $\pi(x^0,\boldsymbol{x})$	
$\varphi(x^0,f) = \int d^3\boldsymbol{x}\, f(\boldsymbol{x})\varphi(x^0,\boldsymbol{x})$ $f \in C_0^\infty(\mathbb{R}^3)$	$\Phi(x^0,f)$, $\Pi(x^0,h)$: operators in $\mathcal{H}$
Approximation $\sum_j q_j(x^0) f(\boldsymbol{x}(j)) a^3 \to \int d^3\boldsymbol{x}\, \varphi(x^0,\boldsymbol{x})$	
$\Rightarrow$ Poisson brackets: $\{\varphi(x^0,f), \pi(x^0,h)\} = \int d^3\boldsymbol{x}\, f(\boldsymbol{x})h(\boldsymbol{x})$	Commutators: $[\Phi(x^0,f), \Pi(x^0,h)] = i\hbar \int d^3\boldsymbol{x}\, f(\boldsymbol{x})h(\boldsymbol{x})$

So far we have introduced field operators $\Phi(x^0,f)$ and their canonically conjugate momenta $\Pi(x^0,h)$, at fixed inertial frame–coordinate time x^0. They are "smeared" against the spatial argument $\boldsymbol{x}$ with test-functions f and h in $C_0(\mathbb{R}^3)$. Without smearing, the density-like quantities $\Phi(x^0,\boldsymbol{x})$ and $\Pi(x^0,\boldsymbol{x})$ cannot be interpreted as operators on a Hilbert space as a consequence of the canonical commutation relations – the entry in the lower right corner of the just tabled

scheme – but only as quadratic forms. This is due to the distributional character of the $\Phi(x^0, \boldsymbol{x})$ and $\Pi(x^0, \boldsymbol{x})$, whence the equal-time canonical commutation relations are often written in the form

$$[\Phi(x^0, \boldsymbol{x}), \Pi(x^0, \boldsymbol{x}')] = i\hbar\delta(\boldsymbol{x} - \boldsymbol{x}')\,.$$

It is quite useful to introduce, for test-functions $F \in C_0^\infty(\mathbb{R}^4)$ distributed over open subsets of Minkowski spacetime, the field operators

$$\Phi(F) = \int d^4x\, F(x^0, \boldsymbol{x})\Phi(x^0, \boldsymbol{x}) = \int dx^0\, \Phi(x^0, f_{x^0})\,, \quad f_{x^0}(\boldsymbol{x}) = F(x^0, \boldsymbol{x})\,.$$

These field operators can be rigorously interpreted as unbounded (and for real-valued F, selfadjoint) operators on a suitable domain of a Hilbert space $\mathcal{H}$ which arises as the bosonic Fock space over the one–particle space of solutions to the Klein–Gordon equation with positive energy. This one–particle space carries an irreducible, unitary representation of $\widetilde{\mathfrak{P}}_+^\uparrow$, which lifts to a unitary representation of $\widetilde{\mathfrak{P}}_+^\uparrow$ on $\mathcal{H}$. Let us denote this representation by U, since it is actually a representation of $\mathfrak{P}_+^\uparrow$ in this case, as for every linear field equation of integer spin. Then one finds that covariance holds in the form of

$$U(L)\Phi(F)U(L)^* = \Phi(F \circ L^{-1})\,, \quad L \in \mathfrak{P}_+^\uparrow, \quad F \in C_0^\infty(\mathbb{R}^4)\,;$$

moreover, one also has

$$\Phi\left(\left(\eta^{\mu\nu}\frac{\partial}{\partial x^\mu}\frac{\partial}{\partial x^\nu} + m^2\right)F\right) = 0\,, \quad F \in C_0^\infty(\mathbb{R}^4)\,, \tag{2}$$

and there holds also the covariant form of the canonical commutation relations,

$$[\Phi(F_1), \Phi(F_2)] = i\hbar G(F_1, F_2)\,, \quad F_1, F_2 \in C_0^\infty(\mathbb{R}^4)\,, \tag{3}$$

with the "causal Green's function"

$$G(F_1, F_2) = \mathrm{Im}\int_{\mathbb{R}^3} \frac{d^3\boldsymbol{p}}{\omega(\boldsymbol{p})}\overline{\tilde{F}_1(\omega(\boldsymbol{p}), -\boldsymbol{p})}\tilde{F}_2(\omega(\boldsymbol{p}), -\boldsymbol{p})$$
$$\omega(\boldsymbol{p}) = \sqrt{\boldsymbol{p}^2 + m^2}\,, \quad \tilde{F} = \text{Fourier-transform of } F$$

which vanishes whenever the supports of F_1 and F_2 are causally separated.

We shall not elaborate on the mathematical details related to the Fock space operators $\Phi(F)$ since this is all well–documented in the literature (see, e.g., [8,49]. Rather we should make the remark at this point that the properties of the operators $\Phi(F)$, interpreted as Fock space operators, may serve as a blue-print of a general concept of a (in this case, scalar) "quantum field", as soon as they are abstracted from properties pertaining to the model of the Klein-Gordon field, i.e., the equation of motion (2). The ensuing conceptual framework for a general scalar quantum field are represented by the "Wightman axioms", which we list now, not paying too much attention to full mathematical rigor (see [8,37,60] for a more detailed exposition of these matters).

i) $\exists$ a Hilbert space $\mathcal{H}$ with a dense domain $\mathcal{D} \subset \mathcal{H}$, so that all $\Phi(F)$ are well-defined operators on $\mathcal{D}$, and $\Phi(F)^* = \Phi(\overline{F})$

ii) $F \mapsto \Phi(F)$ is complex linear and suitably continuous

iii) Covariance: There is on $\mathcal{H}$ a unitary representation $\mathfrak{P}_+^\uparrow \ni L \mapsto U(L)$, with $U(L)\mathcal{D} \subset \mathcal{D}$, so that

$$U(L)\Phi(F)U(L)^* = \Phi(F \circ L^{-1}) \qquad (\ \Phi(x)_L = \Phi(L(x))\ )$$

iv) Locality, or relativistic causality:
If the *supports* of the test-function F_1 and F_2 are *causally separated*, the corresponding field operators commute:

$$[\Phi(F_1), \Phi(F_2)] = 0$$

v) Spectrum condition/positivity of the total energy:
Writing $U(1, a) = e^{iP_\mu a^\mu}$, it holds (in the sense of expectation values) that

$$P_0^2 - P_1^2 - P_2^2 - P_3^2 \geq 0\,, \quad P_0 \geq 0$$

vi) Existence (and uniqueness) of the vacuum:
$\exists\ \Omega \in \mathcal{D}$, $||\Omega|| = 1$, so that $U(L)\Omega = \Omega$ and this vector is uniquely determined up to a phase factor.

vii) Cyclicity of the vacuum:
The domain $\mathcal{D}$ is spanned by vectors of the form

$$\Omega,\ \Phi(F)\Omega,\ \Phi(F_1)\Phi(F_2)\Omega, \ldots,\ \Phi(F_1)\cdots\Phi(F_n)\Omega, \ldots$$

As indicated above, the just given collection of conditions tries to capture the essential properties of a "quantum field". We notice that, compared to the properties of the Klein–Gordon field, the commutation relations (3) have been generalized to the condition of spacelike commutativity, and the reference to a specific field equation has been dropped. Spacelike commutativity says that there should be no uncertainty relations between observables measured at causal separation from each other, and thus gives expression to the principle that there is no operational signal propagation faster than the speed of light. It should be remarked here that there is no difficulty in generalizing the above stated conditions to fields of general spinor- or tensor-type [8, 37, 60]. The basic difference is that for fields of half-integer spin, spacelike commutativity of the field operators must be replaced by spacelike anti-commutativity in order to ensure consistency with the other conditions: This is, basically, the content of the spin-statistics theorem. In this sense, a field carrying half-integer spin does not have the character of an observable – typically, it also transforms non-trivially under gauge transformations. Observable quantities, and related quantum field operators fulfilling spacelike commutation relations, can be built from half–integer spin quantum fields by forming suitable bilinear expressions in those fields. Once more, we must refer to the literature for a fuller discussion of these matters [8].

Furthermore, it is worth noting that the type of Poincaré covariance **iii**), implemented by a unitary representation of $\mathfrak{P}_+^\uparrow$, makes an explicit appearance here, completely in the spirit of the Wigner-Bargmann theorem. (For fields of half–integer spin type, this must be replaced by a unitary representation of $\widetilde{\mathfrak{P}}_+^\uparrow$, in keeping with the circumstance that such fields are not directly observable.)

Some new aspect appears here which we have already alluded to in remark (2.1) and which made an implicit appearance elsewhere when we referred to irreducible unitary representations of $\widetilde{\mathfrak{P}}_+^\uparrow$ having *positive energy*. This is the aspect that the time–translations which the unitary representation $\widetilde{U}$ of $\widetilde{\mathfrak{P}}_+^\uparrow$ implements on the Hilbert space $\mathcal{H}$ are interpreted also as dynamical evolutions of the system, and that these dynamical evolutions be stable in the sense that their corresponding energy is always non-negative and that there should be a common state of lowest energy, the vacuum state. This state is "void of stable particles" but, as we shall see later, not void of correlations, and these have actually a rich structure.

It is the subtle interplay of dynamical stability in the form of the spectrum condition together with locality (or spacelike anti–commutativity in the case of quantum fields carrying half-integer spin) which is responsible for this richness. The condition of cyclicity is mainly made for mathematical convenience; it says that all state vectors of the theory can be approximated by applying polynomials of all field operators on the vacuum. In case of the presence of a vacuum vector, this property could be sharpened to irreducibility, i.e. that already all observables can be approximated by polynomials in the field operators. This is actually equivalent to clustering of vacuum expectation values [8, 37, 60]. However, in a more general situation where there is no vacuum state for all time–evolutions (time-shifts), but e.g., a thermal equilibrium state, irreducibility doesn't hold in general.

While the Wightman framework captures apparently many essential aspects of (observable) quantum fields and is so far not in obvious conflict with experiences gained in constructive quantum field theory, there are some points which lead one to trade this framework for a still more abstract approach. Let me try to illustrate some of these points. The first is of a more technical nature: In handling the – in general – unbounded field operators $\Phi(F)$, subtle domain questions come into play whose physical significance is often not entirely clear. More seriously, it might happen that the field operators $\Phi(F)$ do not correspond to directly observable quantities, and then it is doubtful why they should be regarded as the basic objects of the formal description of a physical theory, at least from an operational point of view. Somehow related to this shortcoming, the $\Phi(F)$ aren't invariants of the experimentally accessible quantities in the following sense: In general, one can find for a given Wightman field $F \mapsto \Phi(F)$ other Wightman fields $F \mapsto \tilde{\Phi}(F)$, subject to different field equations and commutation relations, which yield the same S–matrix as the field $F \mapsto \Phi(F)$ ([10], see also [51] for a more recent instance of this fact). Apart from that, gauge–carrying quantum fields do not fit completely into the framework. Assuming them to be local fields in the same sense as described above often leads to difficulties with Hilbert space

positivity, as e.g. in quantizing free electrodynamics. This difficulty can be cured symptomatically by allowing $\mathcal{H}$ to carry an inner product that is not positive definite [8]. However, such a complication makes technical issues, such as domain questions, even much worse.

Hence, there is considerable motivation to base the description of a relativistic quantum system on observable quantities and to abandon the mainly classically inspired concept of (a quantized version of) a field. In the case that $F \mapsto \Phi(F)$ is an observable quantum field, one can pass to a description of this system which emphasizes the localization of observables in space and time rather than their arrangement into "field strengths": One can form, for each open subset O of Minkowski spacetime $M \equiv \mathbb{R}^4$, a $*$–algebra of bounded operators

$$\begin{aligned}\mathcal{R}(O) = \{&*\text{-algebra generated by all } A = f(\Phi(F))\,,\\ &f : \mathbb{R} \to \mathbb{R} \text{ bounded}\,, \quad F = \overline{F} \text{ has support in } O\}\end{aligned} \tag{4}$$

The properties that one finds for the family of $*$–algebras $\mathcal{R}(O)$, O ranging over the bounded subsets of $\mathbb{R}^4$, form the conditions of the operator algebraic approach to general quantum field theories according to Haag and Kastler [28, 29]. These conditions read as follows.

a) Isotony: $O_1 \subset O_2 \quad \Rightarrow \quad \mathcal{R}(O_1) \subset \mathcal{R}(O_2)$.
b) Covariance: $A \in \mathcal{R}(O) \Leftrightarrow U(L)AU(L)^* \in \mathcal{R}(L(O))$, or $U(L)\mathcal{R}(O)U(L)^* = \mathcal{R}(L(O))$.
c) Locality: If the space–time regions O_1 and O_2 are *causally separated*, then the corresponding operator algebras $\mathcal{R}(O_1)$ and $\mathcal{R}(O_2)$ *commute elementwise*:

$$A \in \mathcal{R}(O_1)\,, \quad B \in \mathcal{R}(O_2) \quad \Rightarrow \quad [A, B] = 0$$

d) Spectrum condition and existence of the vacuum: As before in **v)** and **vi)**, see page 143.
e) Cyclicity of the vacuum: $\{A\Omega : A \in \bigcup_O \mathcal{R}(O)\}$ is dense in $\mathcal{H}$.
f) Weak additivity: If $\bigcup_i O_i$ contains O, then the algebra generated by the $\mathcal{R}(O_i)$ contains $\mathcal{R}(O)$.

We should emphasize that, adopting this framework as basis for a description of a special relativistic quantum system, the crucial structural ingredient is the assignment of not just a single operator algebra to the system but of operator algebras $\mathcal{R}(O)$ to the individual sub-regions O of Minkowski spacetime. Each $\mathcal{R}(O)$ is generated by the observables which can be measured at times and locations in O, and therefore one refers to the observables in $\mathcal{R}(O)$ as those *localized in* O, and to the $\mathcal{R}(O)$ as *local observable algebras*. If actually there is a quantum field $F \mapsto \Phi(F)$ generating the local observable algebras as in (4), then one may view it as a "coordinatization" of the family $\{\mathcal{R}(O)\}_{O \subset M}$, the latter being the "invariant" object, in analogy to a manifold built up from coordinate systems.

A set of data $(\{\mathcal{R}(O)\}_{O \subset M}, U, \Omega)$ fulfilling the conditions just listed is called a *quantum field theory in vacuum representation*. One can consider other representations of a quantum field theory, e.g. thermal representations, where the

spectrum condition imposed on U and the vacuum vector condition imposed on Ω are replaced by the condition that the state $\langle\Omega|\,.\,|\Omega\rangle$ be a thermal equilibrium state. We will encounter such a situation later.

The reader might wonder at this point how charge carrying quantum fields fit into this operator algebraic version of quantum field theory where up to now only observable quantities have been mentioned. The answer is that charge carrying field operators arise in connection with yet other Hilbert space representations of the quantum field field theory, i.e., of the family of operator algebras $\{\mathcal{R}(O)\}_{O\subset M}$. States in these Hilbert space representations cannot be coherently superposed with any state in the vacuum representation. These charged representations are therefore called superselection sectors. The analysis of superselection sectors and the full reconstruction of a compact gauge group and of associated charge carrying quantum field operators from the structure of superselection sectors can be regarded as being one of the greatest achievements in axiomatic quantum field theory so far, but we shall not pause to explore these matters and refer the reader to [21, 28, 54] for further information.

It should be pointed out that all quantum fields obeying linear equations of motion provide examples for the operator algebraic framework, by the relation (4) [for integer spin fields; for half–integer spin fields, one must instead define $\mathcal{R}(O)$ by first constructing suitable bilinear expressions in the fields]. Moreover, there are examples of interacting quantum fields in 2 and 3 spacetime dimensions and these are compatible with the operator algebraic framework via (4).

The interplay between the spectrum condition and locality puts non–trivial constraints on quantum field theories and leads to interesting general results about their structure. Prime examples are the PCT theorem, the spin–statistics relation (cf. [8, 28, 60]) and geometric modular action. About the latter, perhaps less familiar, but highly fascinating issue we have more to report in the following section.

3 The Reeh–Schlieder–Theorem and Geometric Modular Action

In 1961, Helmut Reeh and Siegfried Schlieder showed that the conditions for a quantum field theory of Wightman type, given above, lead to a remarkable consequence [50]. Namely, let O be any non–void open region in Minkowski spacetime, and denote by $\mathcal{P}(O)$ the $*$–algebra generated by all quantum field operators $\Phi(F)$ where the test–functions are supported in O. Then the set of vectors $\mathcal{P}(O)\Omega$, Ω denoting the vacuum vector, is dense in the Hilbert space $\mathcal{H}$. In other words, given an arbitrary vector $\psi\in\mathcal{H}$, and $\epsilon>0$, there is a polynomial

$$\lambda_0\mathbf{1}+\sum_{j,k_j\leq N}\Phi(F_{1,j})\cdots\Phi(F_{k_j,j}) \tag{5}$$

in the field operators, with $\lambda_0\in\mathbb{C}$ and $F_{\ell,j}\in C_0^\infty(O)$, such that

$$||\,\psi-(\lambda_0\mathbf{1}+\sum_{j,k_j\leq N}\Phi(F_{1,j})\cdots\Phi(F_{k_j,j}))\Omega\,||<\epsilon\,. \tag{6}$$

In the operator algebraic setting of Haag and Kastler, the analogous property states that the set of vectors $\mathcal{R}(O)\Omega = \{A\Omega : A \in \mathcal{R}(O)\}$ is dense in $\mathcal{H}$ whenever O is a non-void open set in M; equivalently, given $\psi \in \mathcal{H}$ and $\epsilon > 0$, there is some $A \in \mathcal{R}(O)$ fulfilling

$$|| \psi - A\Omega || < \epsilon \,. \tag{7}$$

This result by Reeh and Schlieder appears entirely counter–intuitive since it says that every state of the theory can be approximated to arbitrary precision by acting with operators (operations) localized in any arbitrarily given spacetime region on the vacuum. To state it in a rather more drastic and provocative way (which I learned from Reinhard Werner): By acting on the vacuum with suitable operations in a terrestrial laboratory, an experimenter can create the Taj Mahal on (or even behind) the Moon!

One might thus be truly concerned that this unusual behavior of relativistic quantum field theory potentially entails superluminal signaling. However, despite the fact that such propositions have been made, this is not the case (see [17,32,56] for some clarifying discussions). We will also turn to aspects of this below in Sect. 4. A crucial point is that the operator $A = A_\epsilon$ in (7) depends on ϵ (and likewise, the polynomial (5) in (6) depends on ϵ), and while $||A_\epsilon\Omega||$ will be bounded (in fact, close to 1) for arbitrarily small ϵ (as follows from (7)), it will in general (in particular, with our drastic Taj Mahal illustration) be the case that A_ϵ doesn't stay bounded as $\epsilon \to 0$, in other words, $||A_\epsilon||$ diverges as ϵ tends to 0.

In keeping with the standard operational interpretation of quantum theory [41], $||A_\epsilon||/||A_\epsilon\Omega||$ is to be viewed as the ratio of cost vs. effect in the attempt to create a given state (Taj Mahal on the Moon) by local operations (in a laboratory on Earth, say) [28]. In other words, upon testing for coincidence with the "Taj Mahal state ψ", it takes on average an ensemble of $||A_\epsilon||/||A_\epsilon\Omega||$ samples failing in the coincidence test to find a single successful coincidence. And in our illustration, the ratio $||A_\epsilon||/||A_\epsilon\Omega||$ will be an enormous number. A rough estimate can be based on the decay of vacuum correlations in quantum field theory. The order of magnitude of that decay is approximately given by $\mathrm{e}^{-d/\lambda_c}$, where d denotes the spatial distance of the correlations and λ_c is the Compton wave length of the stable particles under consideration; then $1/\mathrm{e}^{-d/\lambda_c}$ is a rough measure for $||A_\epsilon||/||A_\epsilon\Omega||$ (when ϵ is very small compared to 1). Taking for instance electrons as stable particles, and the distance Earth–Moon for d, one obtains an order of magnitude of about $10^{-10^{20}}$ for $\mathrm{e}^{-d/\lambda_c}$. This shows that one can hardly construe a contradiction to special relativity on account of the Reeh–Schlieder–theorem.

Nevertheless, for distances that are comparable to the Compton wavelength, the Reeh–Schlieder–theorem does predict a behavior of the correlations in the vacuum state which is in principle experimentally testable, and is of truly quantum nature in the sense that they entail quantum entanglement over subsystems, as will be seen later in Sect. 4.

We will complement the previous discussion by a couple of remarks.

(3.1) The mathematical cause for the Reeh-Schlieder theorem lies in the spectrum condition, which entails that, for each ψ in the Hilbert space of a quantum field theory's vacuum representation, the function

$$(a_1, \ldots, a_n) \mapsto \langle \psi, U(a_1)A_1U(a_2)A_2 \cdots U(a_n)A_n\Omega \rangle, \ A_j \in \mathcal{R}(O), \ a_j \in \mathbb{R}^4, \tag{8}$$

is the continuous boundary value of a function which is analytic in a conical subregion of $\mathbb{C}^{4n}$. Hence, if the expression (8) vanishes when the a_j are in an arbitrarily small open subset of $\mathbb{R}^4$, then it vanishes for all $a_j \in \mathbb{R}^4$. Together with weak additivity one can conclude from this that any vector ψ which is orthogonal to $\mathcal{R}(O)\Omega$ is actually orthogonal to $\bigcup_O \mathcal{R}(O)\Omega$ and hence, by cyclicity of the vacuum vector, ψ must be equal to 0.

(3.2) There are many other state vectors $\xi \in \mathcal{H}$ besides the vacuum vector for which the Reeh-Schlieder theorem holds as well, i.e. for which $\mathcal{R}(O)\xi = \{A\xi : A \in \mathcal{R}(O)\}$ is a dense subset of $\mathcal{H}$ whenever $O \subset M$ is open and non-void. In fact, one can show that there is a dense subset $\mathcal{X}$ of $\mathcal{H}$ so that every $\xi \in \mathcal{X}$ has the property that $\mathcal{R}(O)\xi$ is dense in $\mathcal{H}$ as soon as $O \subset M$ is a non-void open set [20]. Now, every element $\xi \in \mathcal{X}$ (assumed to be normalized) induces a state (expectation value functional)

$$\omega_\xi(A) = \langle \xi, A\xi \rangle, \quad A \in \mathcal{R}(\mathbb{R}^4),$$

and owing to the Reeh-Schlieder property of the vectors $\xi \in \mathcal{X}$, ω_ξ will have long-range correlations, meaning that generically

$$\omega_\xi(AB) \neq \omega_\xi(A)\omega_\xi(B)$$

for $A \in \mathcal{R}(O_\mathrm{A})$ and $B \in \mathcal{R}(O_\mathrm{B})$ even if the spacetime regions O_A and O_B are separated by an arbitrarily large spacelike distance. However, even though the set of vectors ξ inducing such long-range correlations is dense in the set of all state vectors in $\mathcal{H}$, there are in general also very many uncorrelated states. In fact, under very general conditions it could be shown that, as soon as a pair of (finitely extended) spacetime regions O_A and O_B separated by a non-zero spacelike distance is given, together with a pair of vectors ξ_A and ξ_B in $\mathcal{H}$ inducing states ω_{ξ_A} and ω_{ξ_B} on the local observable algebras $\mathcal{R}(O_\mathrm{A})$ and $\mathcal{R}(O_\mathrm{B})$, respectively, there is a state vector $\eta \in \mathcal{H}$ inducing a state ω_η on $\mathcal{R}(\mathbb{R}^4)$ with the property

$$\omega_\eta(AB) = \omega_{\xi_\mathrm{A}}(A)\omega_{\xi_\mathrm{B}}(B), \quad A \in \mathcal{R}(O_\mathrm{A}), \quad B \in \mathcal{R}(O_\mathrm{B}).$$

That is to say, in restriction to the algebra of observables associated to the region $O_\mathrm{A} \cup O_\mathrm{B}$ the state ω_η coincides with the (prescribed) product state induced by the pair of states ω_{ξ_A} and ω_{ξ_B} which has no correlations between the subsystems $\mathcal{R}(O_\mathrm{A})$ and $\mathcal{R}(O_\mathrm{B})$. We should like to refer the reader to [16,64] for considerable discussion on this issue.

(3.3) There are states $\xi \in \mathcal{X}$ for which the Reeh-Schlieder correlations are much stronger that in the vacuum Ω, and in such states the correlations are

sufficiently strong so that they can be used for quantum teleportation over macroscopic distances as has been demonstrated experimentally [26]. While this is perhaps intuitively less surprising than for the case of the vacuum state since the states ξ have some "material content" to which one could ascribe the storage of correlation information, it should be kept in mind that also here the correlations are non-classical, i.e. they manifestly exemplify quantum entanglement.

(3.4) In the Haag-Kastler setting, local commutativity and the Reeh-Schlieder theorem together imply that any local operator $A \in \mathcal{R}(O)$, O open and bounded, which annihilates the vacuum: $A\Omega = 0$, must in fact be equal to the zero operator, $A = 0$. As a consequence, for the vacuum vector Ω (as well as for any other $\xi \in \mathcal{X}$ having the Reeh-Schlieder property) it holds that

$$\langle \Omega, A^* A\Omega \rangle > 0$$

for all $A \in \mathcal{R}(O)$ with $A \neq 0$, O open and bounded. This may be interpreted as the generic presence of vacuum fluctuations; every *local* counting instrument will give a non–zero expectation value in the vacuum state. This is, actually, a situation where relativistic quantum field theory deviates from quantum mechanics. (Quantum mechanics needs to postulate the existence of fluctuations as e.g. in the semiclassical theory of radiation to account for spontaneous emission.)

A related mathematical argument shows that quantities like the energy density will fail to be pointwise positive in the quantum field setting, in contrast to their classical behavior. Yet, the spectrum condition puts limitations to the failure of positivity. For this circle of questions, we recommend that the reader consults the review article [24].

Now, in order to turn to the discussion of "geometric modular action", we need to introduce some notation. We consider a generic *von Neumann algebra* $\mathcal{R}$ acting on a Hilbert space $\mathcal{H}$, together with a unit vector $\Omega \in \mathcal{H}$ which is assumed to be *cyclic and separating* for $\mathcal{R}$. To explain the terminology, $\mathcal{R}$ is a von Neumann algebra acting on $\mathcal{H}$ if $\mathcal{R}$ is a weakly closed (in the sense of convergence of expectation values) $*$-subalgebra of $B(\mathcal{H})$ containing the unit operator. One can show that this is equivalent to the property that $\mathcal{R}$ coincides with its double commutant $\mathcal{R}''$, where the commutant $\mathcal{C}'$ of a subset $\mathcal{C}$ of $B(\mathcal{H})$ is defined as $\mathcal{C}' = \{B \in B(\mathcal{H}) : BC = CB \ \forall\ C \in \mathcal{C}\}$, and the double commutant is then defined by $\mathcal{C}'' = (\mathcal{C}')'$. One says that $\Omega \in \mathcal{H}$ is cyclic for $\mathcal{R}$ if $\mathcal{R}\Omega$ is dense in $\mathcal{H}$ – in view of our previous discussion, this is the same as saying that the Reeh-Schlieder property holds for Ω, with respect to the algebra $\mathcal{R}$. Moreover, one says that Ω is separating for $\mathcal{R}$ if $A \in \mathcal{R}$ and $A\Omega = 0$ imply $A = 0$, and this is equivalent to $\langle \Omega, A^* A\Omega \rangle > 0$ for all $A \in \mathcal{R}$ different from 0. One can in fact show that Ω is cyclic for $\mathcal{R}$ if and only if Ω is separating for $\mathcal{R}'$, and vice versa.

Given a von Neumann algebra $\mathcal{R}$ on a Hilbert space $\mathcal{H}$ and a cyclic and separating unit vector, $\Omega \in \mathcal{H}$, for $\mathcal{R}$, there is a canonical anti-linear operator

$$S : \mathcal{R}\Omega \to \mathcal{R}\Omega\,, \quad A\Omega \mapsto S(A\Omega) := A^*\Omega$$

associated with these data. By cyclicity of Ω for $\mathcal{R}$, the set $\mathcal{R}\Omega = \{A\Omega : A \in \mathcal{R}\}$ is a dense linear subspace of $\mathcal{H}$, so the operator is densely defined; furthermore, to assign the value $A^*\Omega$ to the vector $A\Omega$ in the domain of S is a well–defined procedure in view of the assumption that Ω is separating for $\mathcal{R}$. The anti–linearity of S is then fairly obvious. What is less obvious is the circumstance that the operator S is usually unbounded (provided $\mathcal{H}$ is infinite–dimensional). Nevertheless, one can show that S is a closable operator and thus the closure of S (which we denote here again by S) possesses a polar decomposition, i.e. there is a unique pair of operators J and Δ so that S can be written as

$$S = J\Delta^{1/2}$$

and where $J : \mathcal{H} \to \mathcal{H}$ is anti-linear and fulfills $J^2 = \mathbf{1}$ while $\Delta = S^*S$ is positive (and selfadjoint on a suitable domain, and usually unbounded). This is nothing but the usual polar decompositon of a closable operator, with the slight complication that the operator S is, by definition, anti-linear instead of linear. See, e.g., [11] for further information.

The operators J and Δ are called the *modular conjugation*, and *modular operator*, respectively, corresponding to the pair $\mathcal{R}, \Omega$. Often, J and Δ are also referred to as the *modular objects* of $\mathcal{R}, \Omega$. Their properties have been investigated by the mathematicians Tomita and Takesaki and hence they appear also under the name *Tomita–Takesaki modular objects*. The important properties of J and Δ which were discovered by Tomita and Takesaki (see, e.g., [9, 11, 65] for a full survey of the mathematical statements which we make in what follows) are, first, that the adjoint action of J maps $\mathcal{R}$ onto its commutant $\mathcal{R}'$: $A \in \mathcal{R} \Leftrightarrow JAJ \in \mathcal{R}'$. This is written in shorter notation as $J\mathcal{R}J = \mathcal{R}'$. One also has that $J\Omega = \Omega$. Secondly, since Δ is an invertible non-negative selfadjoint operator, $\ln(\Delta)$ can be defined as a selfadjoint operator by the functional calculus, and hence one can define a one–parametric unitary group $\Delta^{it} = \mathrm{e}^{it\ln(\Delta)}$, $t \in \mathbb{R}$, on $\mathcal{H}$, called the *modular group* of $\mathcal{R}$ and Ω. It has the property that its adjoint action leaves $\mathcal{R}$ invariant, i.e. $A \in \mathcal{R} \Leftrightarrow \Delta^{it}A\Delta^{-it} \in \mathcal{R}$, or simply $\Delta^{it}\mathcal{R}\Delta^{-it} = \mathcal{R}$. Moreover, $\Delta^{it}\Omega = \Omega$ holds for all $t \in \mathbb{R}$. A third property relates to the spectral behavior of the unitary group $\{\Delta^{it}\}_{t\in\mathbb{R}}$. Namely, the state $\omega_\Omega(A) = \langle \Omega, A\Omega\rangle$ on $\mathcal{R}$ fulfills the KMS (Kubo–Martin–Schwinger) boundary condition with respect to the adjoint action of Δ^{it}, $t \in \mathbb{R}$, at inverse temperature $\beta = 1$.

Let us explain the terminology used here. If $\mathcal{R}$ is a von Neumann algebra modeling the observables of a quantum system and $\{\sigma_t\}_{t\in\mathbb{R}}$ is a one–parametric (continuous) group of automorphisms of $\mathcal{R}$ modeling the dynamical evolution of the system, then a density matrix state $\omega_\rho(A) = \mathrm{trace}(\rho A)$ on $\mathcal{R}$ is said to fulfill the KMS boundary condition with respect to $\{\sigma_t\}_{t\in\mathbb{R}}$ (shorter: is a KMS state for $\{\sigma_t\}_{t\in\mathbb{R}}$) at inverse temperature $\beta > 0$ provided that the following holds: Given any pair of elements $A, B \in \mathcal{R}$, there exists a function F_{AB} which is analytic on the complex strip $\mathcal{S}_\beta = \{t + i\eta : t \in \mathbb{R},\ 0 < \eta < \beta\}$, and is continuous on the closure of the strip $\mathcal{S}_\beta$, with the boundary values

$$F_{AB}(t) = \omega_\rho(\sigma_t(A)B)\,, \quad F_{AB}(t + i\beta) = \omega_\rho(B\sigma_t(A))\,, \qquad t \in \mathbb{R}\,.$$

For a quantum mechanical system with a Hamilton operator H such that $\mathrm{e}^{-\beta H}$ is a trace–class operator ($\beta > 0$), one can form the density matrices

$$\rho^{\beta} = \frac{1}{\mathrm{trace}(\mathrm{e}^{-\beta H})}\mathrm{e}^{-\beta H}$$

and one can show that the corresponding Gibbs states ω_{ρ^β} are KMS states at inverse temperature β for the dynamical evolution given by $\sigma_t(A) = \mathrm{e}^{itH}A\mathrm{e}^{-itH}$. Haag, Hugenholtz and Winnink [30] have shown that states of an infinite quantum system – being modeled by $\mathcal{R}$ and $\{\sigma_t\}_{t\in\mathbb{R}}$ – which are suitably approximated by Gibbs states of finite subsystems, are under very general conditions also KMS states, and thus the KMS boundary condition is viewed as being characteristic of thermal equilibrium states.

Therefore, if ω_Ω is a KMS state with respect to the (adjoint action of the) modular group $\{\Delta^{it}\}_{t\in\mathbb{R}}$ of $\mathcal{R}$,Ω, this signalizes that there is some relation to physics provided that $\{\Delta^{it}\}_{t\in\mathbb{R}}$ can be interpreted as dynamical evolution of a quantum system. This is not always the case, but the converse always holds true: Suppose that a quantum system dynamical system consisting of $\mathcal{R}$ and $\{\sigma_t\}_{t\in\mathbb{R}}$ and a KMS state ω_ρ at inverse temperature $\beta > 0$ is given. Then one can pass to the *GNS (Gelfand–Naimark–Segal) representation* associated with $\mathcal{R}$ and ω_ρ. This is a triple $(\pi^\rho, \mathcal{H}^\rho, \Omega^\rho)$ where $\mathcal{H}^\rho$ is a Hilbert space, π^ρ is a representation of $\mathcal{R}$ by bounded linear operators on $\mathcal{H}^\rho$ (which may differ from the "defining" representation of $\mathcal{R}$ that is pre-given since the elements of $\mathcal{R}$ act as bounded linear operators on a Hilbert space $\mathcal{H}$) and Ω^ρ is a unit vector in $\mathcal{H}^\rho$ which is cyclic for $\pi^\rho(\mathcal{R})$ and with $\omega^\rho(A) = \langle\Omega^\rho, \pi^\rho(A)\Omega^\rho\rangle$. In this GNS representation, $\{\sigma_t\}_{t\in\mathbb{R}}$ is implemented by the (rescaled) modular group $\{\Delta^{it/\beta}\}_{t\in\mathbb{R}}$ corresponding to $\pi^\rho(\mathcal{R})''$ and Ω^ρ: $\pi^\rho(\sigma_t(A)) = \Delta^{it/\beta}\pi^\rho(A)\Delta^{-it/\beta}$.

Tomita–Takesaki theory has had a considerable impact on the development of operator algebra theory. Owing to its relation to thermal equilibrium states, it has also found applications in quantum statistical mechanics. It took longer, however, until a connection between Tomita–Takesaki modular objects and the action of the Poincaré group was revealed in the context of relativistic quantum field theory. Such a connection was established in the seminal work of Bisognano and Wichmann [7]. To explain their result, let (x^0, x^1, x^2, x^3) denote the coordinates of points in Minkowski spacetime in some Lorentzian frame. Then let $W = \{x = (x^0, x^1, x^2, x^3) \in \mathbb{R}^4 : x^1 > 0,\ -x^1 < x^0 < x^1\}$ denote the "right wedge region" with respect to the chosen coordinates. Moreover, we shall introduce the following maps of Minkowski spacetime:

$$\mathbf{j} : (x^0, x^1, x^2, x^3) \mapsto (-x^0, -x^1, x^2, x^3)$$

which is a reflection about the spatial x^2-x^3 plane together with a time-reflection, and

$$\Lambda_1(\theta) = \begin{pmatrix} \cosh(\theta) & -\sinh(\theta) & 0 & 0 \\ -\sinh(\theta) & \cosh(\theta) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \theta \in \mathbb{R},$$

the Lorentz boosts along the x^1-axis, which map W onto itself.

Now consider a quantum field theory of the Haag–Kastler type (in vacuum representation), where it is also assumed that the local algebras of observables $\mathcal{R}(O)$ are generated by a Wightman-type quantum field $F \mapsto \Phi(F)$ as in (4). It will also be assumed that the $\mathcal{R}(O)$ are actually von Neumann algebras, so that one has $\mathcal{R}(O) = \mathcal{R}(O)''$ for open, bounded regions. Then one can also built the algebra of observables located in the wedge region W,

$$\mathcal{R}(W) = \{A \in \mathcal{R}(O) : O \subset W\}'' .$$

We will denote by J the modular conjugation and by $\{\Delta^{it}\}_{t\in\mathbb{R}}$ the modular group, respectively, associated with $\mathcal{R}(W)$ and the vacuum vector Ω. These are well–defined since the vacuum vector is, by the Reeh–Schlieder–theorem, cyclic and separating for $\mathcal{R}(W)$. With these assumptions, Bisognano and Wichmann [7] found the following remarkable result.

Theorem The following relations hold:

$$\begin{aligned}
\Delta^{it} \quad &= U(\Lambda_1(2\pi t)) \\
J\mathcal{R}(O)J &= R(\mathbf{j}(O)) , \quad \text{moreover,} \\
J\Phi(F)J &= \Phi(\overline{F \circ \mathbf{j}}) , \\
JU(L)J &= U(\mathbf{j} \circ L \circ \mathbf{j}) , \quad L \in \mathfrak{P}_+^\uparrow
\end{aligned}$$

Here, U denotes the unitary representation of the Poincaré group belonging to the quantum field theory under consideration, and we have written $U(\Lambda_1(2\pi t))$ for the unitary representation of the Lorentz boost $\Lambda_1(2\pi t)$.

The remarkable point is that by this theorem, the modular conjugation and modular group associated with $\mathcal{R}(W)$ and Ω acquire a clear–cut geometric meaning. Moreover, since the adjoint action of J involves, in its geometric meaning, a time and space reflection, it induces a PCT symmetry in the following way:

The rotation $D_{(2,3)}$ by $\pi = 180°$ in the (x^2, x^3) plane is contained in the proper, orthochronous Poincaré group, and

$$\mathbf{j} \circ D_{(2,3)} = D_{(2,3)} \circ \mathbf{j} = PT : x \mapsto -x$$

is the total inversion.

Then $\Theta = JU(D_{(2,3)})$ is a *PCT operator*: Θ is anti-unitary and fulfills $\Theta^2 = \mathbf{1}$, and

$$\begin{aligned}
\Theta\Omega \quad &= \Omega \\
\Theta\mathcal{R}(O)\Theta &= \mathcal{R}(PT(O)) \\
\Theta\Phi(F)\Theta &= \Phi(\overline{F \circ PT}) \\
\Theta U(L)\Theta &= U(PT \circ L \circ PT) .
\end{aligned}$$

Because of the geometric significance of the modular objects J and $\{\Delta^{it}\}_{t\in\mathbb{R}}$ one also says that the Bisognano–Wichmann theorem is an instance of "geometric

modular action" (although this term is actually used also in a wider context). The concept of "geometric modular action" has been used quite fruitfully in the analysis of general quantum field theories over the past years and has led to remarkable progress and insights. We cannot get into this matter in any depth and instead we refer the reader to the comprehensive review by Borchers [9]; we will only comment on a few aspects of geometric modular action by way of a couple of remarks.

(3.5) Because of $\Delta^{it} = U(\Lambda_1(2\pi t))$, the vacuum state functional $\langle \Omega, \, . \, \Omega \rangle$ restricted to $\mathcal{R}(W)$ is a KMS state, i.e. a thermal equilibrium state. More precisely, an observer following the trajectory

$$\gamma_a(t) = \Lambda_1(t) \begin{pmatrix} 0 \\ 1/a \\ 0 \\ 0 \end{pmatrix}$$

will register the (restriction of the) vacuum state along his or her trajectory as a thermal equilibrium state at absolute temperature

$$T_a = \frac{\hbar a}{2\pi k c} \,,$$

where here we have explicitly inserted $\hbar$, Boltzmann's constant k and the velocity of light c. This is called the *Fulling-Unruh-effect* [25, 66]. It has been noted by Sewell [58, 61] that a similar form of geometric modular action for quantum fields on the Schwarzschild-Kruskal spacetime can be viewed as a variant of the Hawking effect.

(3.6) The relation of Tomita–Takesaki objects to the action of the Poincaré group which is displayed by the Bisognano-Wichmann theorem is only realized if the observable algebras with respect to which the Tomita-Takesaki objects those belonging to wedge regions – i.e. any Poincaré-transform of W. For observable algebras $\mathcal{R}(O)$ belonging to bounded regions, the corresponding modular objects have in general no clear geometric meaning. An exception is the case of a conformal quantum field theory when O is a double cone (see [9] and literature cited there).

(3.7) If a is a lightlike vector parallel to the future lightlike boundary of W, let

$$J_a = \text{modular conjugation of } \; \mathcal{R}(W + a), \Omega$$

Then one can show that

$$U(-2a) = J_0 J_a \,,$$

i.e. the modular conjugations encode the translation group – together with the spectrum condition. Since the modular group of $\mathcal{R}(W)$ induces the boosts, it appears that the complete unitary action of the Poincaré group can be retrieved from the modular objects of observable algebras

belonging to a couple of wedge regions in suitable position to each other, with a common vacuum vector. And indeed, a careful analysis has shown that this is possible under general conditions [9, 38]. This opens the possibility to approach the problem of constructing (interacting) quantum field theories in a completely novel manner, where one starts with a couple of von Neumann algebras together with a common cyclic and separating vector, and where the associated modular objects fulfil suitable relations so that they induce a representation of the Poincaré group. See [14, 57] for perspectives, first steps and results around this circle of ideas.

(3.8) It should also be pointed out that geometric modular action can be understood in a more general sense than above where the modular objects associated with the vacuum and algebras of observables located in wedge-regions induce point-transformations on the manifold – in our present discussion, always Minkowski spacetime – on which the quantum field theory under consideration lives. A more general criterion of geometric modular action would, e.g., be the following: Given a family of observable (von Neumann) algebras $\{R(O)\}_{O\subset M}$ indexed by the open (and bounded) subsets of a spacetime manifold M, and a vector Ω in the Hilbert space representation of that family, one can try to find a sub-family $\{R(\tilde{O})\}_{\tilde{O}\in\tilde{K}}$ (where $\tilde{K}$ is a collection of subsets of M, sufficiently large so that a base of the topology of M can be generated by countable intersections and unions of members in $\tilde{K}$, say) with the property that the adjoint action of the modular conjugation $J_{\hat{O}}$ of $\mathcal{R}(\hat{O}), \Omega$, where $\hat{O}$ is any element of $\tilde{K}$, maps the family $\{R(\tilde{O})\}_{\tilde{O}\in\tilde{K}}$ onto itself. This would be a generalized form of geometric modular action. In the light of the Bisognano–Wichmann theorem, for the case of Minkowski spacetime one would take the collection of wedge regions as $\tilde{K}$ and the vacuum vector as Ω. But there are instances where precisely such a generalized form of geometric modular action is realized when taking for M e.g. Robertson-Walker spacetimes. For more discussion on this intriguing generalization of geometric modular action, see [15].

4 Relativistic Quantum Information Theory: Distillability in Quantum Field Theory

The final section of this contribution is devoted to a subject which seems to be of growing interest nowadays [4, 22, 47, 52]: The attempt to bring together the flourishing discipline of quantum information theory with the principles of special relativity. Since quantum information theory is based on the principles of quantum mechanics and since quantum field theory is the theory which unifies quantum mechanics and special relativity, it appears entirely natural to discuss issues of relativistic quantum information theory in the setting of quantum field theory.

There are, of course, several foundational issues one might wish to discuss when studying a prospective merging of quantum information theory and special relativity even in the established setting of quantum field theory. One of them might be the so far omitted discussion on quantum measurement theory within quantum field theory. In view of the Reeh–Schlieder–theorem, one may suspect delicate problems at this point – in fact, there are numerous discussions on the nature of locality/nonlocality in quantum (information) theory, where sometimes the various authors don't agree on precisely what sort of locality is attributed to which object or structure within a particular theoretical framework. Our approach here is operational, and we refer to works already cited [32, 56] for some discussion on measurement in quantum field theory.

This said, we limit ourselves here to studying a very particular concept which has been developed and investigated in non-relativistic quantum information theory in the context of relativistic quantum field theory: The concept of *distillability* of quantum states. Very roughly speaking, one can say that distillable quantum states contain "useful" entanglement that can be enhanced, at least theoretically, so that it can be used as a resource for typical telecommunication tasks such as quantum cryptography or quantum teleportation [5, 23, 26]. (For a more detailed exposition of the formal apparatus of quantum information theory and important references, we recommend the review by M. Keyl [40].) To make this more precise, we will now have to specify our setting at a more formal level. Everything what follows is taken from a joint publication with R. Werner [67].

First, we will say that a *bipartite system* is a pair of mutually commuting $*$-subalgebras $\mathcal{A}$, $\mathcal{B}$ of $B(\mathcal{H})$ for some Hilbert space $\mathcal{H}$. Usually, we will in fact assume that both $\mathcal{A}$ and $\mathcal{B}$ are von Neumann algebras; one could also generalize the setting by only requiring that $\mathcal{A}$ and $\mathcal{B}$ are $*$-subalgebras of a common C^*-algebra.

In the quantum field theoretical context, $\mathcal{A}$ will be identified with $\mathcal{R}(O_{\mathrm{A}})$ and $\mathcal{B}$ with $\mathcal{R}(O_{\mathrm{B}})$ for a pair of (bounded) spacetime regions O_{A} and O_{B} which are causally separated. Quite generally, $\mathcal{A}$ represents the algebra of observables in a laboratory controlled by a physicist named 'Alice' and $\mathcal{B}$ represents the algebra of observables in a laboratory controlled by another physicist called 'Bob'. The prototypical example of a bipartite system in (non-relativistic) quantum information theory is the situation where $\mathcal{H} = \mathcal{H}_{\mathrm{A}} \otimes \mathcal{H}_{\mathrm{B}}$, and where $\mathcal{A} = B(\mathcal{H}_{\mathrm{A}}) \otimes \mathbf{1}$ and where $\mathcal{B} = \mathbf{1} \otimes B(\mathcal{H}_{\mathrm{B}})$. The situation in relativistic quantum field theory can be a bit more complicated.

Let $\mathcal{A}, \mathcal{B} \subset B(\mathcal{H})$ form a bipartite quantum system, and let $\omega(X) = \mathrm{trace}(\rho X)$, for some density matrix ρ on $\mathcal{H}$, be a state on $B(\mathcal{H})$. We say that the state ω is a *product state* on the bipartite system if $\omega(AB) = \omega(A)\omega(B)$ holds for all $A \in \mathcal{A}$ and all $B \in \mathcal{B}$. Moreover, ω is called *separable* on the bipartite system if it is a limit (in the sense of convergence of expectation values) of convex combinations of product states. Then, ω is called *entangled* on the bipartite system if it is not separable.

Entanglement of states on bipartite systems is a typical quantum phenomenon with no counterpart in classical physics. As is well known, the

Einstein–Podolsky–Rosen paradoxon really centers about entangled states, as has been clarified and formalized by John Bell (see the reprint collection [69] for the relevant references and comments, and the textbook [46] for a more modern and simpler discussion). As mentioned, nowadays entanglement is viewed as a resource for tasks of quantum communication, and this circumstance has motivated several studies on the "degree" or "quality" of entanglement that a state may have (see, again, the review [40] for discussion and references). One possible measure of "entanglement strength" is provided by the *Bell-CHSH correlation* [19, 63]. This is a number, $\beta(\omega)$, which is assigned to any state ω of a bipartite system $\mathcal{A}, \mathcal{B} \subset B(\mathcal{H})$ as

$$\beta(\omega) = \sup_{A,A',B,B'} \omega(A(B'+B) + A'(B'-B))$$

where the supremum is taken over all hermitean $A, A' \in \mathcal{A}$ and $B, B' \in \mathcal{B}$ whose operator norm is bounded by 1. Separable states always have $\beta(\omega) \leq 2$. This case is referred to by saying that ω *fulfills the Bell-CHSH inequalities*. States ω for which $\beta(\omega) > 2$ are said to *violate the Bell-CHSH inequalities*; such states are entangled. The maximal number which $\beta(\omega)$ can assume is $2\sqrt{2}$ [18], and states for which $\beta(\omega) = 2\sqrt{2}$ are said to *violate the Bell-CHSH inequalities maximally*. In a sense, one may view a state ω_1 more strongly entangled than a state ω_2 if $\beta(\omega_1) > \beta(\omega_2)$.

Let us consider a particularly simple system where $\mathcal{H} = \mathbb{C}^2 \otimes \mathbb{C}^2$, with $\mathcal{A} = B(\mathbb{C}^2) \otimes \mathbf{1}$ and $\mathcal{B} = \mathbf{1} \otimes B(\mathbb{C}^2)$, where $B(\mathbb{C}^2)$ is a perhaps slightly unusual way to denote the algebra of complex 2×2 matrices. A state violating the Bell–CHSH inequalities maximally is given by the singlet state $\omega_{\text{singlet}}(X) = \langle \psi_{\text{singlet}}, X\psi_{\text{singlet}} \rangle$, $X \in B(\mathbb{C}^2 \otimes \mathbb{C}^2)$, where

$$\psi_{\text{singlet}} = \frac{1}{\sqrt{2}}(|0\rangle \otimes |1\rangle - |1\rangle \otimes |0\rangle)\,;$$

here, $|0\rangle$ and $|1\rangle$ denote the two orthonormalized eigenvectors of the Pauli–matrix σ_z. There are, in fact, experimental situations in quantum optics where the singlet state can be realized to a high degree of accuracy. In these situations, one identifies $|0\rangle$ and $|1\rangle$ with the two orthonormal polarization states of photons which are linearly polarized with respect to chosen coordinates perpendicular to the direction of propagation. One can prepare a source (state) producing an ensemble of pairs of polarized photons in the singlet state and send – e.g. through optical fibers over long distances – one member of each ensemble pair to the laboratory of Alice (whose observables, regarding the polarization of the photons, are represented by $\mathcal{A}$) and the other member of the same pair to the laboratory of Bob (whose polarization observables are represented by $\mathcal{B}$). In this way, Alice and Bob have access to a common entangled state ω_{singlet} which they may use for carrying out tasks of quantum communication. The singlet state (or rather, any singlet–type state) is, in this sense, the best suited state owing to its "maximal" entanglement which is reflected by its maximal violation of the Bell-CHSH inequalities. Some experimental realizations and applications can be found e.g. in [26].

There are entangled states ω which are not as strongly entangled as ω_{singlet}, but contain still enough entanglement so that a sub-ensemble of photon pairs can be "distilled" from ω which coincides with ω_{singlet} to high accuracy and may then be used for carrying out quantum communication tasks. To make such a "distillability" an attribute of the given state ω, one must ensure that the distillation process only enhances the entanglement already present in the given state ω, and doesn't induce previously non–existing entanglement. One tries to capture this requirement by demanding that the process of distillation involves only local operations and classical communication (LOCC) [6, 40, 48].

The idea behind LOCC is best illustrated by a simple example. We assume that both Alice and Bob operate a two–valued instrument in each of their laboratories. A two–valued instrument (i) takes an incoming state, (ii) puts out either of two classical values ("readings"), say "+" or "−" and (iii) changes the state into a new output state depending on the values of the classical readings, i.e. the values "+" or "−". Thus, if the source (represented by the state ω) produces a pair of polarized photons, then the pair member running to Alice passes her instrument while the other pair member travels to Bob and passes his instrument. The pair members are then subjected to state changes – operations – taking place individually at the sites of the laboratories of Alice and Bob, respectively, and are thus local (assuming that the operations are active at mutual spacelike separation); put differently, Alice's instrument operates only on the pair member in her laboratory and likewise Bob's instrument operates only on the pair member in his laboratory. We further suppose that Alice and Bob agree to discard all photon pairs except those which on passing their instruments have yielded in both cases the "+" reading. Since they don't now beforehand what the values of these readings will be, they have to inform each other about the readings' values of their instruments *after* both members of each photon pair have passed through. This requires "two-way classical communication" between Alice and Bob. Then, after a large number of photon pairs (corresponding, in idealization, to the original ensemble of the state ω) has passed the instruments, and having discarded all the pairs not giving the "+" reading, Alice and Bob hold (in each lab, members of) a smaller number (a subensemble) of photon pairs which have been subjected to local operations mediated by the instruments. This new subensemble may correspond to a state with stronger entanglement, and if, in this way, a subensemble can be produced which approximates the singlet state ω_{singlet} to arbitrary precision, then the original state ω is called *distillable*. Strictly speaking, we should call the state *1-distillable*, the qualifier "1" referring to only "1 round" of instrument application and classical communication for each photon pair, since one can envisage more complicated schemes of using localized (multi-valued) instruments and classical communication between Alice and Bob that are still in compliance with the idea of local operations and classical communication. But then, any state which is 1–distillable will also be distillable according to a more general scheme, so that 1–distillability is in this sense the most stringent criterion.

Now we need to give a mathematical description of 1–distillability of a state ω. In the present simple case, the mathematical image of a two-valued instrument in Alice's laboratory is given by two completely positive maps $T_\pm : \mathcal{A} \to \mathcal{A}$ with $T_+(\mathbf{1}) + T_-(\mathbf{1}) = \mathbf{1}$. Likewise, in Bob's laboratory, his two-valued instrument is given by a pair of completely positive maps $S_\pm : \mathcal{B} \to \mathcal{B}$ with $S_+(\mathbf{1}) + S_-(\mathbf{1}) = \mathbf{1}$. The subensemble that Alice and Bob select from the original state ω corresponds to the positive functional $\mathcal{A} \otimes \mathcal{B} \ni x \otimes y \mapsto \omega(T_+(x)S_+(y))$, which is turned into a state, $\mathcal{A} \otimes \mathcal{B} \ni x \otimes y \mapsto \omega(T_+(x)S_+(y))/\omega(T_+(\mathbf{1})S_+(\mathbf{1}))$, upon normalization. Let us denote this new state by $\omega^{T,S}$, identifying T with T_+ and S with S_+. To say that ω is 1-distillable now amounts to requiring that one can choose S and T in such a way that $\omega^{T,S}$ approximates $\omega_{\mathrm{singlet}}$ to arbitrary precision.

All this applies as yet to the case that $\mathcal{A}$ and $\mathcal{B}$ are copies of $B(\mathbb{C}^2)$. However, it is not too difficult to generalize everything to the case of a generic bipartite quantum system. All that needs to be done is to ensure that the input state ω, defined on the algebra generated by $\mathcal{A}$ and $\mathcal{B}$, yields an output state $\omega^{T,S}$ on $B(\mathbb{C}^2 \otimes \mathbb{C}^2)$ which can be compared to $\omega_{\mathrm{singlet}}$. The formal definition of 1-distillability is then:

Definition Let ω be a state on a general bipartite quantum system $\mathcal{A}, \mathcal{B} \subset B(\mathcal{H})$. The state ω is called *1-distillable* if one can find completely positive maps $T : B(\mathbb{C}^2) \to \mathcal{A}$ and $S : B(\mathbb{C}^2) \to \mathcal{B}$ so that the state

$$\omega^{T,S}(x \otimes y) = \omega(T(x)S(y))/\omega(T(\mathbf{1})S(\mathbf{1}))\,, \quad x \otimes y \in B(\mathbb{C}^2 \otimes \mathbb{C}^2)\,,$$

on $B(\mathbb{C}^2 \otimes \mathbb{C}^2)$ approximates $\omega_{\mathrm{singlet}}$ to arbitrary precision. That is to say, given $\epsilon > 0$, there are such $T = T_\epsilon$ and $S = S_\epsilon$ so that

$$|\,\omega^{T,S}(X) - \omega_{\mathrm{singlet}}(X)\,| < \epsilon||X||\,, \quad X \in B(\mathbb{C}^2 \otimes \mathbb{C}^2)\,. \tag{9}$$

This criterion for 1–distillability is now completely general and can, in particular, be applied in the context of relativistic quantum field theory. This is what we will do now.

As in Sect. 2, let $(\{\mathcal{R}(O)\}_{O \subset M}, U, \Omega)$ be a quantum field theory in vacuum representation. We quote following result, taken from [67].

Theorem Let $\mathcal{A} = \mathcal{R}(O_{\mathrm{A}})$ and $\mathcal{B} = \mathcal{R}(O_{\mathrm{B}})$ be a bipartite quantum system formed by algebras of local observables localized in spacetime regions O_{A} and O_{B} which are separated by a non-zero spacelike distance. Then the vacuum state $\omega(\,.\,) = \langle \Omega, \,.\,\Omega\rangle$ is 1-distillable on this bipartite system. Moreover, there is a dense set $\mathcal{X} \subset \mathcal{H}$ so that the vector states $\omega_\chi(\,.\,) = \langle \chi, \,.\,\chi\rangle$, $||\chi|| = 1$, are 1-distillable on the bipartite system. (In fact, $\mathcal{X}$ can be chosen so that this holds for all spacelike separated regions O_{A} and O_{B}.)

We will add a couple of remarks.

(4.1) The conclusion of the theorem remains valid if one considers the quantum field theory in a relativistic thermal equilibrium representation instead of a vacuum representation. Representations of this kind have been introduced by Bros and Buchholz [13]. The distinction from the vacuum prepresentation is as follows: The spectrum condition is dropped, and it is assumed

that $\omega(\,.\,) = \langle\Omega,\, .\,\Omega\rangle$ fulfills the *relativistic KMS condition* at some inverse temperature $\beta > 0$. Following [13], one says that a state ω on $\mathcal{R}(\mathbb{R}^4)$ satisfies the relativistic KMS condition at inverse temperature $\beta > 0$ (with respect to the adjoint action of the translation group $U(a)$, $a \in \mathbb{R}^4$) if there exists a timelike vector e in V_+, the open forward light cone, so that e has unit Minkowskian length, and so that for each pair of operators $A, B \in \mathcal{R}(\mathbb{R}^4)$ there is a function $F = F_{AB}$ which is analytic in the domain $\mathcal{T}_{\beta e} = \{z \in \mathbb{C}^4 : \mathrm{Im}\, z \in V_+ \cap (\beta e - V_+)\}$, and continuous at the boundary sets determined by $\mathrm{Im}\, z = 0$, $\mathrm{Im}\, z = \beta e$ with the boundary values $F(a) = \langle\Omega, AU(a)B\Omega\rangle$, $F(a + i\beta e) = \langle\Omega, BU(-a)A\Omega\rangle$ for $a \in \mathbb{R}^4$. Upon comparison with the non-relativistic KMS–condition of the previous section, one may get an idea in which way this is a relativistic generalization of thermal equilibrium states.

(4.2) It is the Reeh-Schlieder theorem which is responsible for the distillability result; we briefly sketch the argument. In fact, one can show that each non-abelian von Neumann algebra contains an isomorphic copy of $B(\mathbb{C}^2)$. In the particular case considered in the situation of Theorem (4.3), one can use the Reeh-Schlieder theorem to prove that there are algebraic morphisms $\tau : B(\mathbb{C}^2) \to \mathcal{A}$ and $\sigma : B(\mathbb{C}^2) \to \mathcal{B}$ so that $\pi : B(\mathbb{C}^2 \otimes \mathbb{C}^2) \to B(\mathcal{H})$ given $\pi(x \otimes y) = \tau(x)\sigma(y)$ is a faithful algebraic embedding. Then there is a unit vector χ in $\mathcal{H}$ so that $\omega_\chi(\pi(X)) = \omega_{\mathrm{singlet}}(X)$ for all $X \in B(\mathbb{C}^2 \otimes \mathbb{C}^2)$. According to the Reeh-Schlieder theorem, there is for any $\epsilon > 0$ some $A = A_\epsilon \in \mathcal{A}$ with $||A|| = 1$ so that $||(||A\Omega||)^{-1}A\Omega - \chi|| < \epsilon$. Thus we choose $T(x) = A^*\tau(x)A$ and $S(y) = \sigma(y)$ to obtain that the state $\omega^{T,S}$ fulfills the required estimate (9). The Reeh-Schlieder theorem for relativistic thermal equilibrium states has been proved in [35].

(4.3) The normalization factor $\omega(T(\mathbf{1})S(\mathbf{1}))$ equals, in the previous remark, the quantity $\langle\Omega, A^*\tau(\mathbf{1})\sigma(\mathbf{1})A\Omega\rangle$ which, in turn, is equal to $||A\Omega||^2$ up to a term of at most order ϵ. Since we have taken $||A||$ to be equal to 1 (which made the occurrence of the normalization factor $(||A\Omega||)^{-1}$ necessary in the approximation of χ), the quantity $||A\Omega||$ here coincides in fact with $||A\Omega||/||A||$, i.e. the effect vs. cost ratio which made its appearance in our discussion of the Reeh-Schlieder theorem. Thus, the factor $\omega(T(\mathbf{1})S(\mathbf{1}))$ (compared to 1) is a rough measure for the efficiency of the distillation process, or put differently, the fraction of members in the subensemble corresponding to $\omega^{T,S}$ distilled from the members of the original ensemble ω. As we have seen before, this will be a very small number when ϵ is small and the spatial distance between the regions O_{A} and O_{B} is macroscopic for ω the vacuum state.

(4.4) We should like to mention that there are many related works addressing the issue of long–range correlations in quantum field theory. In fact, Bell–correlations in quantum field theory have been investigated before quantum information theory was established; see the refs. [42, 43, 62, 63], and they have contributed to understand quantum entanglement in a mathematically rigorous form applicable to general quantum systems.

More recent works in this direction prove that in the bipartite situation $\mathcal{A} = \mathcal{R}(O_{\rm A})$, $\mathcal{B} = \mathcal{R}(O_{\rm B})$, for a relativistic quantum field theory, there is dense set of states violating the Bell-CHSH inequalities [31,36,52]. In this sense, they are quite closely related to the result of the theorem above, which however gives also information about the distillability of specific states, such as the vacuum or relativistic thermal equilibrium states, over arbitrarily spacelike subsystems of a relativistic quantum field theory.

I think that, in the light of the theoretical developments summarized in this contribution, it is fair to say that the interplay between special relativity and quantum physics is holding a significant position at the frontier of current research. Thus I am quite confident that special relativity will live well through the next 101 years.

References

1. Araki, H, *General principles of quantum field theory*, Oxford University Press, 1990
2. Bargmann, V., Ann. Math. **49** (1947) 568 & Ann. Math. **59** (1954) 1; J. Math. Phys. **5** (1964) 862
3. Baumgärtel, H., *Operatoralgebraic methods in quantum field theory*, Akademie Verlag, Berlin, 1995
4. Beckman, D., Gottesman, D., Nielsen, M.A., Preskill, J., Phys. Rev. **A 64** (2001) 052309
5. Bennett, C.H., Brassard, G., Crépeau, C., Josza, R., Peres, A., Wootters, W.K., Phys. Rev. Lett. **70** (1993) 1; Bennett, C.H., Brassard, G., Popescu, S., Schumacher, B., Smolin, J.A., Phys. Rev. Lett. **76** (1996) 722; erratum ibid. **78** (1997) 2031; Gisin, N., Phys. Lett A **210** (1996) 151
6. Bennett, C.H., DiVincenzo, D.P., Smolin, J.A., Wootters, W.K., Phys.Rev.A **54** (1996) 3824
7. Bisognano, J.J., Wichmann, E.H., J. Math. Phys. **16** (1975) 985 & **17** (1976) 303
8. Bogoliubov N.N., Lugonov, A.A., Oksak A.I., Todorov, I.T., *General principles of quantum field theory*, Kluwer, Dordrecht, 1990
9. Borchers, H.-J., J. Math. Phys. **41** (2000) 3604
10. Borchers, H.-J., Nuovo Cimento **15** (1960) 784
11. Bratteli, O., Robinson, D.W., *Operator algebras and quantum statistical mechanics*, Vol. 1, 2nd ed., Springer, Berlin-Heidelberg-New York, 1987
12. Bratteli, O., Robinson, D.W., *Operator algebras and quantum statistical mechanics*, Vol. 2, 2nd ed., Springer, Berlin-Heidelberg-New York, 1997
13. Bros, J., Buchholz, D., Nucl. Phys. **B429**, 291 (1994)
14. Buchholz, D., Lechner, G., Ann. H. Poincaré **5** (2004) 1065
15. Buchholz, D., Mund, J., Summers, S.J., Fields Inst. Commun. **30** (2001) 65
16. Buchholz, D., Wichmann, E.H., Commun. Math. Phys. **106** (1986) 321
17. Buchholz, D., Yngvason, J., Phys. Rev. Lett. **73** (1994) 613
18. Cirel'son, B.S., Lett. Math. Phys. **4**, 93 (1980).
19. Clauser, J.F., Horne, M.A., Shimony, A., Holt, R.A., Phys. Rev. Lett. **26** (1969) 880 (1969)
20. Dixmier, J., Maréchal, O., Commun. Math. Phys. **22** (1971) 44
21. Doplicher, S., Roberts, J.E., Commun. Math. Phys. **131** (1990) 51

22. Eggeling, T., Schlingemann, D., Werner, R.F., Europhys. Lett. **57**, 782 (2001)
23. Ekert, A.K., Phys. Rev. Lett. **67**, 661 (1991)
24. Fewster, C.J., "Energy inequalities in quantum field theory", math-ph/0501073
25. Fulling, S.A., Phys. Rev. **D7** (1973) 2850
26. Gisin, N., Ribordy, G., Tittel, W., Zbinden, H., Rev. Mod Phys. **74** (2002) 145; A. Poppe et al., "Practical quantum key distribution with polarization entangled photons", quant-ph/0404115; Jennewein, T., Ursin, R., Aspelmeyer, M., Zeilinger, A., "Experimental teleportation of quantum entanglement with an optimal linear optics Bell-state analyzer", quant-ph/0409008; Jennewein, T., Weihs, G., Pan, J.-W., Zeilinger, A., Phys. Rev. Lett **88** (2002) 017903; Weihs, G., Jennewein, T., Simon, C., Weinfurter, H., Zeilinger, A., Phys. Rev. Lett **81** (1998) 5039
27. Glimm, J., Jaffe, A., *Quantum physics*, Springer-Verlag, New York, 1987
28. Haag, R., *Local quantum physics*, Springer-Verlag, Berlin, 1992
29. Haag, R., Kastler, D., J. Math. Phys. **5** (1964) 848
30. Haag, R., Hugenholtz, N.M., Winnink, M., Commun. Math. Phys. **5** (1967) 215
31. Halvorson, H., Clifton, R., J. Math. Phys. **41** (2000) 1711
32. Hellwig, K.-E., Kraus, K., Phys. Rev. **D1** (1970) 566 & Commun. Math. Phys. **16** (1970) 142
33. Henley, E.M., Thirring, W., *Elementary quantum field theory*, McGraw-Hill, New York, 1962
34. Itzykson, C., Zuber, *Quantum field theory*, McGraw-Hill, Singapore, 1985
35. Jäkel, C.D., J. Math. Phys. **41** (2000) 1745
36. Jäkel, C.D., Found. Phys. Lett. **14** (2001) 1
37. Jost, R., *The general theory of quantum fields*, AMS, Providence, 1965
38. Kähler, R., Wiesbrock, H.-W., J. Math. Phys. **42** (2001) 74
39. Keyl, M., Schlingemann, D., Werner, R.F., Quant. Inf. Comp. **3** (2003) 281
40. Keyl, M., Phys. Rep. **369** (2002) 431
41. Kraus, K., *States, effects and operations*, Springer Lecture Notes in Physics, Springer-Verlag, Berlin, 1972
42. Landau, L.J., Phys. Lett. **A120** (1987) 54
43. Landau, L.J., Phys. Lett. **A123** (1987) 115
44. Montvay, I., Münster, G., *Quantum fields on a lattice*, Cambridge University Press, 1994
45. Pais, A., *Subtle is the Lord. The science and life of Albert Einstein*, Oxford University Press, 1982
46. Peres, A., *Quantum theory: concepts and methods*, Kluwer, Dordrecht, 1993
47. Peres, A., Terno, D.R., Rev. Mod. Phys. **76** (2004) 93
48. Popescu, S., Phys. Rev. Lett. **74** (1995) 2619
49. Reed, C., Simon, B., *Methods of mathematical physics*, Vol. II, Academic Press, New York, 1975
50. Reeh, H., Schlieder, S., Nuovo Cimento **22** (1961) 1051
51. Rehren, K.-H., Commun. Math. Phys. **178** (1996) 453
52. Reznik, B., Retzker, A., Silman, J., Phys. Lett. **A337** (2005) 17
53. Rivasseau, V., *From perturbative to constructive renormalization*, Princeton University Press, 1991
54. Roberts, J.E., *More lectures on algebraic quantum field theory*, in: S. Doplicher and R. Longo (Eds.), Springer Lecture Notes in Mathematics 1831, pp. 263-342, Springer-Verlag, Berlin 2004
55. Roberts, J.E., Roepstorff, G., Commun. Math. Phys. **11** (1968) 321
56. Schlieder, S., Commun. Math. Phys. **7** (1968) 305

57. Schroer, B., Wiesbrock, H.-W., Rev. Math. Phys. **12** (2000) 301
58. Sewell, G.L., Ann. Phys. **141** (1982) 201
59. Sexl, R.U., Urbantke, H.K., *Relativity, groups, particles*, Springer-Verlag, Wien, 2001
60. Streater, R.F., Wightman, A.S., *PCT, spin and statistics, and all that*, Benjamin, New York, 1964
61. Summers, S.J., Verch, R., Lett. Math. Phys, **37** (1996) 145
62. Summers S.J., Werner, R.F., Lett. Math. Phys. **33** (1995) 321 ; – Ann. Inst. H. Poincaré **49** (1988) 215; – Commun. Math. Phys. **110** (1987) 247; – Phys. Lett. A **110** (1985) 257
63. Summers, S.J., Werner, R.F., J. Math. Phys. **28** (1987) 2448
64. Summers, S.J., Rev. Math. Phys. **2** (1990) 201
65. Takesaki, M., *Tomita's theory of modular Hilbert algebras and its applications* Springer Lecture Notes in Mathematics 128, Springer-Verlag, Berlin, 1970
66. Unruh, W.G., Phys. Rev. **D14** (1976) 870
67. Verch, R., Werner, R.F., Rev. Math. Phys. **17** (2005) 545-576
68. Weinberg, S., *The quantum theory of fields*, Vols. I-III, Cambridge University Press, 1996-2000
69. Wheeler, J.A., Zurek, W.H. (Eds.), *Quantum theory and measurement*, Princeton University Press, 1983
70. Wigner, E.P., Ann. math. **40** (1939) 149

Spacetime Metric from Local and Linear Electrodynamics: A New Axiomatic Scheme

F.W. Hehl[1] and Y.N. Obukhov[2]

[1] Inst. Theor. Physics, University of Cologne, 50923 Köln, Germany
Dept. Physics Astron., Univ. of Missouri-Columbia, Columbia, MO 65211, USA
hehl@thp.uni-koeln.de
[2] Inst. Theor. Physics, University of Cologne, 50923 Köln, Germany
Dept. Theor. Physics, Moscow State University 117234 Moscow, Russia
yo@thp.uni-koeln.de

Abstract. We consider *spacetime* to be a 4-dimensional differentiable manifold that can be split locally into time and space. No metric, no linear connection are assumed. *Matter* is described by classical fields/fluids. We distinguish electrically charged from neutral matter. *Electric charge* and *magnetic flux* are postulated to be conserved. As a consequence, the inhomogeneous and the homogeneous Maxwell equations emerge expressed in terms of the excitation $H = (\mathcal{H}, \mathcal{D})$ and the field strength $F = (E, B)$, respectively. H and F are assumed to fulfill a *local and linear* "spacetime relation" with 36 constitutive functions. The propagation of electromagnetic waves is considered under such circumstances in the geometric optics limit. We forbid *birefringence* in vacuum and find the light cone including its Lorentzian signature. Thus the conformally invariant part of the metric is recovered. If one sets a scale, one finds the pseudo-Riemannian metric of spacetime.

1 Introduction

The neutrinos, in the standard model of elementary particle physics, are assumed to be massless. By the discovery of the neutrino oscillations, this assumption became invalidated. The neutrinos are massive, even though they carry, as compared to the electron, only very small masses. Then the *photon* is left as the only known massless and free elementary particle. The gluons do not qualify in this context since they are confined and cannot exist as free particles under normal circumstances.

Consequently, the photon is the only particle that is directly related to the light cone $g_{ij}dx^i \otimes dx^j = 0$ and that can be used for an operational definition of the light cone; here g_{ij} is the metric of spacetime, dx^i a coordinate differential, and $i, j = 0, 1, 2, 3$. We are back – as the name light cone suggests anyway – to an electromagnetic view of the light cone. Speaking in the framework of classical optics, the *light ray* would then be the elementary object with the help of which one can span the light cone. We take "light ray" as synonymous for radar signals,

F.W. Hehl and Y.N. Obukhov: *Spacetime Metric from Local and Linear Electrodynamics: A New Axiomatic Scheme*, Lect. Notes Phys. **702**, 163–187 (2006)
DOI 10.1007/3-540-34523-X_7 © Springer-Verlag Berlin Heidelberg 2006

laser beams, or electromagnetic rays of other wavelengths. It is understood that classical optics is a limiting case, for short wavelengths, of classical Maxwell–Lorentz electrodynamics.

In other words, if we assume the framework of Maxwell–Lorentz electrodynamics, we can derive, in the geometric optics limit, light rays and thus the light cone, see Perlick [48] and the literature given there. However, the formalism of Maxwell–Lorentz electrodynamics is interwoven with the Riemannian metric g_{ij} in a nontrivial way. Accordingly, in the way sketched, one can never hope to find a real derivation of the light cone.

Therefore, we start from the *premetric* form of electrodynamics, that is, a metric of spacetime is not assumed. Nevertheless, we can derive the generally covariant Maxwell equations, expressed in terms of the excitation $H = (\mathcal{H}, \mathcal{D})$ and the field strength $F = (E, B)$, from the conservation laws of electric charge and magnetic flux. We assume a local and linear spacetime relation between H and F. Then we can solve the Maxwell equations. In particular, we can study the propagation of electromagnetic waves, and we can consider the geometrical optics limit. In this way, we derive the light rays that are spanning the light cone. In general, we find a *quartic* wave covector surface (similar as in a crystal) that only reduces to the pseudo-Riemannian light cone of general relativity if we forbid *birefringence* (double refraction) in vacuum. Hence, in the framework of premetric electrodynamics, the local and linear spacetime relation, together with a ban on birefringence in vacuum, yields the pseudo-Riemannian light cone of general relativity. Accordingly, the geometrical structure of a Riemannian spacetime is derived from purely electromagnetic data. We consider that as our contribution to the Einstein year 2005, and we hope that going beyond the geometrical optics limit will yield better insight into the geometry of spacetime.

The axiomatic scheme that we are going to present here is already contained in our book [20] where also references to earlier work and more details can be found. In the meantime we learnt from the literature that appeared since 2003 (see, e.g., Delphenich [8, 9], Itin [24], Kaiser [26], Kiehn [29], and Lindell & Sihvola [35, 37]) and improved our derivation of the light cone, simplified it, made it more transparent (see, e.g., [21, 23, 33, 45, 46]). The formalism and the conventions we take from [20].

2 Spacetime

In our approach, we start from a 4-dimensional spacetime manifold that is just a continuum which can be decomposed locally in (1-dimensional) time and (3-dimensional) space. It carries no metric and it carries no (linear or affine) connection. As such it is inhomogeneous. It doesn't make sense to assume that a vector field is constant in this continuum. Only the constancy of a scalar field is uniquely defined. Also a measurement of temporal or spatial intervals is still not defined since a metric is not yet available.

In technical terms, the spacetime is a 4-dimensional connected, Hausdorff, paracompact, and oriented differential manifold. On such a manifold, we assume

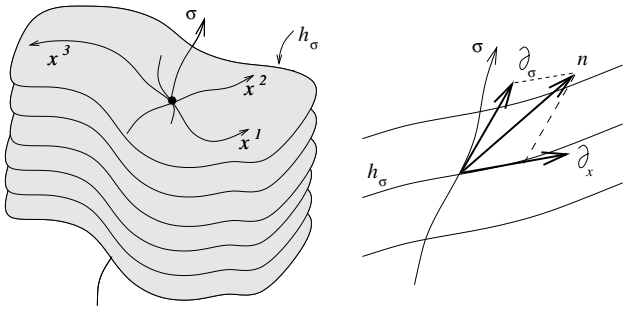


Fig. 1. Local spacetime foliation, see [20]

the existence of a foliation: The spacetime can be decomposed locally into three-dimensional folios labeled consecutively by a monotonically increasing "prototime" parameter σ, see Fig. 1. A vector field n, transverse to the foliation, is normalized by $n\rfloor d\sigma = \mathcal{L}_n\sigma = 1$. Accordingly, we find for the dimensions $[n] = [\sigma]^{-1} = t^{-1}$, where t denotes the dimension of time.

We can decompose any exterior form Ψ in "time" and "space" pieces. The part *longitudinal* to the vector n reads

$$^{\perp}\Psi := d\sigma \wedge \Psi_{\perp}\,, \qquad \Psi_{\perp} := n\rfloor\Psi\,, \tag{1}$$

the part *transversal* to the vector n

$$\underline{\Psi} := (1 \; - \; ^{\perp})\Psi = n\rfloor(d\sigma \wedge \Psi)\,, \quad n\rfloor\,\underline{\Psi} \equiv 0\,. \tag{2}$$

Putting these two parts together, we have the space-time decomposition

$$\Psi = \; ^{\perp}\Psi + \underline{\Psi} = d\sigma \wedge \Psi_{\perp} + \underline{\Psi}\,, \tag{3}$$

with the absolute dimensions $[\Psi_{\perp}] = [\Psi]\,t^{-1}$ and $[\underline{\Psi}] = [\Psi]$.

The 3-dimensional exterior derivative is defined by $\underline{d} := n\rfloor(d\sigma \wedge d)$. We can use the notion of the *Lie derivative* of a p-form Ψ along a vector field ξ, i.e., $\mathcal{L}_\xi\Psi := \xi\rfloor d\Psi + d(\xi\rfloor\Psi)$, and can introduce the derivative of a transversal field $\underline{\Psi}$ with respect to prototime as

$$\dot{\underline{\Psi}} := \mathcal{L}_n\underline{\Psi}\,. \tag{4}$$

3 Matter – Electrically Charged and Neutral

We assume that spacetime is "populated" with classical matter, either described by fields and/or by fluids. In between the agglomerations of matter, there may also exist vacuum.

Matter is divided into electrically charged and neutral matter. Turning to the physics of the former, we assume that on the spatial folios of the manifold we can determine an electric charge Q as a 3-dimensional integral over a charge density and a magnetic flux Φ as a 2-dimensional integral over a flux density.

This is at the bottom of classical electrodynamics: Spacetime is filled with matter that is characterized by charge Q and by magnetic flux Φ. For neutral matter both vanish. The absolute dimension of charge will be denoted by q, that of magnetic flux by $\phi = [\text{action}/\text{charge}] = h/q$, with h as the dimension of action.

4 Electric Charge Conservation

One can catch single electrons and single protons in traps and can *count* them individually. Thus, the electric charge conservation is a fundamental law of nature, valid in macro- as well as in micro-physics.[3] Accordingly, it is justified to introduce the absolute dimension of charge q as a new and independent concept.

Let us define, in 4-dimensional spacetime, the electrical current 3-form J, with dimension $[J] = q$. Its integral over an arbitrary 3-dimensional spacetime domain yields the total charge contained therein: $Q = \int_{\Omega_3} J$. Accordingly, the local form of charge conservation (Axiom 1) reads:

$$d\,J = 0\,. \tag{5}$$

This law is metric-independent since it is based on a *counting* procedure for the elementary charges. Using a foliation of spacetime, we can decompose the current J into the 2-form of the electric current density j and the 3-form ρ of the electric charge density:

$$J = -j \wedge d\sigma + \rho\,. \tag{6}$$

Then (5) can be rewritten as the continuity equation:

$$\dot{\rho} + \underline{d}\, j = 0. \tag{7}$$

Both versions of charge conservation, (5) and (7), can also be formulated in an integral form.

5 Charge Active: Excitation

Electric charge was postulated to be conserved in all regions of spacetime. If spacetime is topologically sufficiently trivial, we find, as consequence of (5), that J has to be exact:

$$J = d\,H\,. \tag{8}$$

This is the inhomogeneous Maxwell equation in its premetric form. The *electromagnetic excitation* 2-form H, with $[H] = [J] = q$, is measurable with the

[3] Lämmerzahl, Macias, and Müller [34] proposed an extension of Maxwell's equations that violates electric charge conservation. Such a model can be used as a test theory for experiments that check the validity of charge conservation, and it allows to give a numerical bound.

help of ideal conductors and superconductors and thus has a direct *operational* significance.

By decomposing H into time and space, we obtain the electric excitation 2-form $\mathcal{D}$ (historical name: "dielectric displacement") and the magnetic excitation 1-form $\mathcal{H}$ ("magnetic field"):

$$H = -\mathcal{H} \wedge d\sigma + \mathcal{D}\,. \tag{9}$$

Substituting (9) into (8), we recover the pair of the 3-dimensional inhomogeneous Maxwell equations

$$d\,H = J \quad \begin{cases} \underline{d}\,\mathcal{D} = \rho\,, \\ -\dot{\mathcal{D}} + \underline{d}\,\mathcal{H} = j\,. \end{cases} \tag{10}$$

6 Charge Passive: Field Strength

With the derivation of the inhomogeneous Maxwell equations the information contained in Axiom 1 is exhausted. As is evident from the Coulomb-Gauss law $\underline{d}\,\mathcal{D} = \rho$, it is the active character of ρ that plays a role in this inhomogeneous Maxwell equation: The charge density ρ is the source of $\mathcal{D}$ (and, analogously, the current density j that of $\mathcal{H}$).

Since we search for new input, it is near at hand to turn to the *passive* character of charge, that is, to wonder what happens when a test charge is put in an electromagnetic field. In the purely electric case with a test charge e, we have

$$F_a \sim e\,E_a\,, \tag{11}$$

with F_a and E_a as components of the covectors of force and electric field strength, respectively. The simplest relativistic generalization for defining the electromagnetic field is then of the type

$$\text{force density} \sim \text{field strength} \times \text{charge current density}\,. \tag{12}$$

Accordingly, with the force density covector (or 1-form) f_α, we can formulate Axiom 2 as

$$f_\alpha = (e_\alpha \rfloor F) \wedge J\,. \tag{13}$$

Here e_α is a local frame, with $\alpha = 0, 1, 2, 3$. Axiom 2 provides an operational definition of the electromagnetic field strength 2-form F, the absolute dimension of which turns out to be $[F] = h/q$. Its $1 + 3$ decomposition

$$F = E \wedge d\sigma + B\,, \tag{14}$$

introduces the electric field strength 1-form E and the magnetic field strength 2-form B, see Fig. 2. If we substitute (14) and (6) into (13), we recover, for $\alpha = 1, 2, 3$, the Lorentz force density.

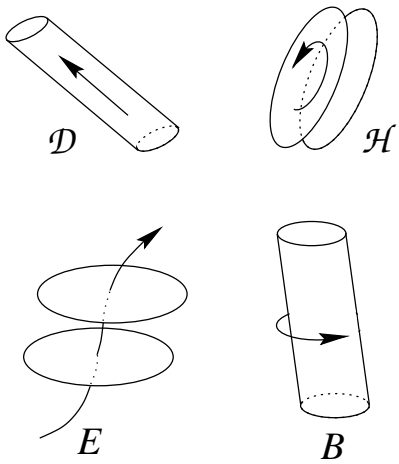


Fig. 2. Faraday-Schouten pictograms of the electromagnetic field, see [20]. The electric excitation $\mathcal{D}$ is a twisted 2-form, the magnetic excitation $\mathcal{H}$ a twisted 1-form. The electric field strength E is a 1-form and the magnetic field strength B a 2-form, both without twist

7 Magnetic Flux Conservation

The field strength F, as a 2-form, can be integrated over a 2-dimensional area Ω_2 in 4-dimensional spacetime. This yields the total magnetic flux Φ piercing through this area: $\Phi = \int_{\Omega_2} F$. In close analogy to electric charge conservation, we assume that also the flux is conserved. Then, in local form, magnetic flux conservation (Axiom 3) reads[4]

$$d\,F = 0 \quad \begin{cases} \underline{d}\,B &= 0\,, \\ \dot{B} + \underline{d}\,E &= 0\,. \end{cases} \tag{15}$$

The Faraday induction law and the sourcelessness of B are the two consequences of $dF = 0$. In this sense, Axiom 3 has a firm experimental underpinning.

8 Premetric Electrodynamics

...is meant to be the "naked" or "featureless" spacetime manifold, without metric and without connection, together with the Maxwell equations $dH = J$,

[4] One can give up magnetic flux conservation by introducing magnetic monopoles according to $dF = J_{\rm magn}$. In premetric electrodynamics this has been done by Edelen [11], Kaiser [26], and by us [21]. However, then one has to change Axiom 2, too, and the Lorentz force density picks up an additional term $-(e_\alpha \rfloor H) \wedge J_{\rm magn}$. This destroys Axiom 2 as an operational procedure for defining F. Moreover, magnetic charges have never been found.

$dF = 0$, the Lorentz force formula, and the electromagnetic energy-momentum current to be discussed below, see (20). We stress that the Poincaré group and special relativity have nothing to do with the foundations of electrodynamics as understood here in the sense of the decisive importance of the underlying *generally covariant* conservation laws of charge (Axiom 1) and flux (Axiom 3). Historically, special relativity emerged in the context of an analysis of the electrodynamics of moving bodies [13, 39], but within the last 100 years classical electrodynamics had a development of its own and its structure is now much better understood than it was 100 years ago. *Diffeomorphism invariance* was recognized to be of overwhelming importance. Poincaré invariance turned out to play a secondary role only.

Of course, premetric electrodynamics so far does not represent a complete physical theory. The excitation H does not yet communicate with the field strength F. Only by specifying a "spacetime" relation between H and F (the constitutive law of the spacetime manifold), only thereby we recover — under suitable conditions — our normal Riemannian or Minkowskian spacetime which we seem to live in. In this sense, a realistic spacetime — and thus an appropriate *geometry* thereof — emerges only by specifying additionally a suitable spacetime relation on the featureless spacetime.

As explained, Axiom 1, Axiom 2, Axiom 3, together with Axiom 4 on energy-momentum, constitute premetric electrodynamics. Let us display the first three axions here again, but now Axiom 1 and Axiom 3 in in the more general integral version. For any submanifolds C_3 and C_2 that are closed, i.e., $\partial C_3 = 0$ and $\partial C_2 = 0$, the axioms read

$$\oint_{C_3} J = 0\,, \qquad f_\alpha = (e_\alpha \rfloor F) \wedge J\,, \qquad \oint_{C_2} F = 0\,. \tag{16}$$

By de Rham's theorem we find the corresponding differential versions

$$d\,J = 0\,, \qquad f_\alpha = (e_\alpha \rfloor F) \wedge J\,, \qquad dF = 0\,, \tag{17}$$

$$J \;\; = d\,H\,, \qquad\qquad\qquad\qquad F = dA\,. \tag{18}$$

The physical interpretation of the quantities involved is revealed via their (1+3)-decompositions (6), (9), (14), and $A = -\varphi d\sigma + \mathcal{A}$, see Fig. 3.

Let us now turn to the energy-momentum question. Using the properties of the exterior differential, we can rewrite the Lorentz force density (13) as

$$f_\alpha = (e_\alpha \rfloor F) \wedge J = d\,{}^{\mathrm{k}}\Sigma_\alpha + X_\alpha\,. \tag{19}$$

Here the *kinematic energy-momentum* 3-form of the electromagnetic field, a central result in the premetric electrodynamics, reads (Axiom 4)

$${}^{\mathrm{k}}\Sigma_\alpha := \frac{1}{2}\left[F \wedge (e_\alpha \rfloor H) - H \wedge (e_\alpha \rfloor F)\right]\,. \tag{20}$$

The remaining force density 4-form turns out to be

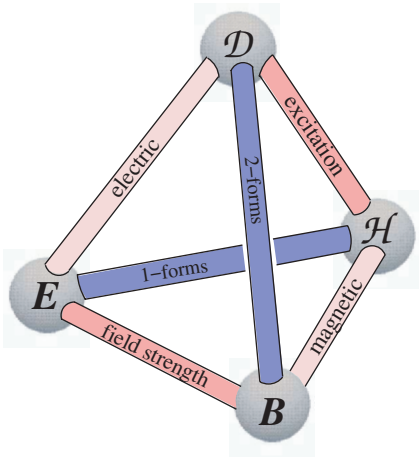


Fig. 3. Different aspects of the electromagnetic field. The four quantities $\mathcal{H}, \mathcal{D}, E, B$ constitute the electromagnetic field. The excitations $\mathcal{H}, \mathcal{D}$ are *extensive* quantities (how much?), the field strengths E, B *intensive* quantities (how strong?)

$$X_\alpha := -\frac{1}{2}\left(F \wedge \mathcal{L}_{e_\alpha} H - H \wedge \mathcal{L}_{e_\alpha} F\right) . \tag{21}$$

The absolute dimension of ${}^{\mathrm{k}}\Sigma_\alpha$ and of X_α is h/ℓ, where ℓ denotes the dimension of length. [Provided, additionally, a linear connection is given with the covariant differential D, then

$$f_\alpha = D\,{}^{\mathrm{k}}\Sigma_\alpha + \widehat{X}_\alpha \,, \tag{22}$$

with the new supplementary force density

$$\widehat{X}_\alpha = \frac{1}{2}\left(H \wedge \mathrm{Ł}_{e_\alpha} F - F \wedge \mathrm{Ł}_{e_\alpha} H\right) , \tag{23}$$

which contains the covariant Lie derivative. In general relativity theory, $\widehat{X}_\alpha$ eventually vanishes for the standard Maxwell–Lorentz electrodynamics.]

9 The Excitation is Local and Linear in the Field Strength

The system of the Maxwell equations $dH = J\,,\ dF = 0$ is apparently underdetermined. It gets predictive power only when we supplement it with a spacetime (or constitutive) relation between the excitation and the field strength. As Axiom 5, we postulate a general local and linear spacetime relation

$$H = \kappa(F)\,, \qquad H_{ij} = \frac{1}{2}\,\kappa_{ij}{}^{kl}\,F_{kl}\,. \tag{24}$$

Here excitation and field strength decompose according to $H = H_{ij}\,dx^i \wedge dx^j/2$ and $F = F_{ij}\,dx^i \wedge dx^j/2$, respectively. The constitutive tensor κ, as 4th rank

tensor with 36 independent components, has to be space and time dependent since constant components would not have a generally covariant meaning on the naked spacetime manifold we consider.

Let us decompose ${\kappa_{ij}}^{kl}$ into irreducible pieces. In the premetric framework we can only perform a *contraction*. A first contraction yields

$$ {\kappa_i}^k := {\kappa_{il}}^{kl} \quad \text{(16 independent functions)}\,, \tag{25} $$

a second one

$$ \kappa := {\kappa_k}^k = {\kappa_{kl}}^{kl} \quad \text{(1 pseudo-scalar function)}\,. \tag{26} $$

Then, introducing the traceless piece

$$ \not{\kappa}_i{}^k := {\kappa_i}^k - \frac{1}{4}\,\kappa\,\delta_i^k \quad \text{(15 functions)}\,, \tag{27} $$

we can rewrite the original constitutive tensor as

$$ \begin{aligned} {\kappa_{ij}}^{kl} &= {}^{(1)}{\kappa_{ij}}^{kl} + {}^{(2)}{\kappa_{ij}}^{kl} + {}^{(3)}{\kappa_{ij}}^{kl} \\ &= {}^{(1)}{\kappa_{ij}}^{kl} + 2\,\not{\kappa}_{[i}{}^{[k}\,\delta_{j]}^{l]} + \frac{1}{6}\,\kappa\,\delta_{[i}^k\delta_{j]}^l. \end{aligned} \tag{28} $$

The *skewon* and the *axion* fields are conventionally defined by

$$ \mathcal{S}_i{}^j = -\frac{1}{2}\,\not{\kappa}_i{}^j, \qquad \alpha = \frac{1}{12}\,\kappa. \tag{29} $$

Substituting (28) into (24) and using (29), we obtain the spacetime relation explicitly:

$$ H_{ij} = \frac{1}{2}\,{}^{(1)}{\kappa_{ij}}^{kl}\,F_{kl} + 2\,\mathcal{S}_{[i}{}^k F_{j]k} + \alpha\,F_{ij}\,. \tag{30} $$

The *principal* (or the metric-dilaton) part ${}^{(1)}{\kappa_{ij}}^{kl}$ of the constitutive tensor with 20 independent components will eventually be expressed in terms of the metric (thereby cutting the 20 components in half). [In standard Maxwell–Lorentz electrodynamics

$$ {}^{(1)}{\kappa_{ij}}^{kl} = \lambda_0\,\sqrt{-g}\,\hat{\epsilon}_{ijmn}\,g^{mk}g^{nl}\,, \quad \mathcal{S}_i{}^k = 0\,, \quad \alpha = 0.] \tag{31} $$

The principal part ${}^{(1)}{\kappa_{ij}}^{kl}$ must be non-vanishing in order to allow for electromagnetic wave propagation in the geometrical optics limit, see the next section. The skewon part $\mathcal{S}_i{}^k$ with its 15 components was proposed by us. We put forward the hypothesis that such a field exists in nature. Finally, the axion part α had already been introduced in elementary particle physics in a different connection but with the same result for electrodynamics, see, e.g., Wilczek's axion electrodynamics [65] and the references given there.

The spacetime relation we are discussing here is the constitutive relation for spacetime, i.e., for the vacuum. However, one has analogous structures for a *medium* described by a local and linear constitutive law. The *skewon* piece in

this framework corresponds to chiral properties of the medium inducing optical activity, see Lindell et al. [38], whereas the concept of an *axion* piece has been introduced by Tellegen [57, 58] for a general medium, by Dzyaloshinskii [10] specifically for Cr_2O_3, and by Lindell & Sihvola [37] in the form of the so-called perfect electromagnetic conductor (PEMC). Recently, Lindell [36] discussed the properties of a *skewon-axion* medium.

The following alternative representation of the constitutive tensor is useful in many derivations and for a comparison with literature, see Post [50],

$$\chi^{ijkl} := \frac{1}{2}\,\epsilon^{ijmn}\,\kappa_{mn}{}^{kl}\,, \tag{32}$$

with

$$\chi^{ijkl} = \underbrace{{}^{(1)}\chi^{ijkl}}_{20,\,\text{principal}} + \underbrace{\epsilon^{ijm[k}\,\mathcal{S}_m{}^{l]} - \epsilon^{klm[i}\,\mathcal{S}_m{}^{j]}}_{15,\,\text{skewon}} + \underbrace{\epsilon^{ijkl}\,\alpha}_{1,\,\text{axion}}\,. \tag{33}$$

It is convenient to consider the excitation H and the field strength F as 6-vectors, each comprising a pair of two 3-vectors. The spacetime relation then reads

$$\begin{pmatrix} \mathcal{H}_a \\ \mathcal{D}^a \end{pmatrix} = \begin{pmatrix} \mathfrak{C}^b{}_a & \mathfrak{B}_{ba} \\ \mathfrak{A}^{ba} & \mathfrak{D}_b{}^a \end{pmatrix} \begin{pmatrix} -E_b \\ B^b \end{pmatrix}. \tag{34}$$

Accordingly, the constitutive tensors are represented by the 6×6 matrices

$$\kappa_I{}^K = \begin{pmatrix} \mathfrak{C}^b{}_a & \mathfrak{B}_{ba} \\ \mathfrak{A}^{ba} & \mathfrak{D}_b{}^a \end{pmatrix}, \qquad \chi^{IK} = \begin{pmatrix} \mathfrak{B}_{ab} & \mathfrak{D}_a{}^b \\ \mathfrak{C}^a{}_b & \mathfrak{A}^{ab} \end{pmatrix}. \tag{35}$$

The 3×3 matrices $\mathfrak{A}, \mathfrak{B}, \mathfrak{C}, \mathfrak{D}$ are defined by

$$\mathfrak{A}^{ba} := \chi^{0a0b}\,, \qquad \mathfrak{B}_{ba} := \frac{1}{4}\,\hat{\epsilon}_{acd}\,\hat{\epsilon}_{bef}\,\chi^{cdef}\,, \tag{36}$$

$$\mathfrak{C}^a{}_b := \frac{1}{2}\,\hat{\epsilon}_{bcd}\,\chi^{cd0a}\,, \qquad \mathfrak{D}_a{}^b := \frac{1}{2}\,\hat{\epsilon}_{acd}\,\chi^{0bcd}\,, \tag{37}$$

or explicitly, recalling the irreducible decomposition (33),

$$\mathfrak{A}^{ab} = -\varepsilon^{ab} - \epsilon^{abc}\;\mathcal{S}_c{}^0, \qquad \mathfrak{B}_{ab} = \mu^{-1}_{ab} + \hat{\epsilon}_{abc}\;\mathcal{S}_0{}^c, \tag{38}$$

$$\mathfrak{C}^a{}_b = \gamma^a{}_b - (\mathcal{S}_b{}^a - \delta^a_b\;\mathcal{S}_c{}^c) + \alpha\,\delta^a_b, \tag{39}$$

$$\mathfrak{D}_a{}^b = \gamma^b{}_a + (\mathcal{S}_a{}^b - \delta^b_a\;\mathcal{S}_c{}^c) + \alpha\,\delta^b_a. \tag{40}$$

The constituents of the principal part are the *permittivity* tensor $\varepsilon^{ab} = \varepsilon^{ba}$, the *impermeability* tensor $\mu^{-1}_{ab} = \mu^{-1}_{ba}$, and the magnetoelectric cross-term $\gamma^a{}_b$, with $\gamma^c{}_c = 0$ (Fresnel-Fizeau effect). The skewon $\mathcal{S}_b{}^a$ and the axion α describe electric and magnetic Faraday effects and (in the last two relations) optical activity. If we substitute (38), (39), (40) into (34), we find a 3-dimensional explicit form of our Axiom 5 formulated in (24):

$$\mathcal{D}^a = \left(\varepsilon^{ab} - \epsilon^{abc}\;\mathcal{S}_c{}^0\right) E_b + \left(\;\gamma^a{}_b + \mathcal{S}_b{}^a - \delta^a_b\;\mathcal{S}_c{}^c\right) B^b + \alpha B^a\,, \tag{41}$$

$$\mathcal{H}_a = \left(\mu^{-1}_{ab} - \hat{\epsilon}_{abc}\;\mathcal{S}_0{}^c\right) B^b + \left(-\gamma^b{}_a + \mathcal{S}_a{}^b - \delta^b_a\;\mathcal{S}_c{}^c\right) E_b - \alpha\,E_a\,. \tag{42}$$

10 Propagation of Electromagnetic Rays ("Light")

After the spacetime relation (Axiom 5) has been formulated, we have a complete set of equations describing the electromagnetic field. We can now study the propagation of electromagnetic waves à la Hadamard. The sourceless Maxwell equations read

$$dH = 0\,, \quad dF = 0\,. \tag{43}$$

In the geometric optics approximation (equivalently, in the Hadamard approach) an electromagnetic wave is described by the propagation of a discontinuity of the electromagnetic field. The surface of discontinuity S is defined locally by a function Φ such that $\Phi = const$ on S. The jumps $[\ ]$ of the electromagnetic quantities across S and the wave covector $q := d\Phi$ then satisfy the geometric Hadamard conditions:

$$[H] = 0\,,\ [dH] = q \wedge h = 0 \qquad \Rightarrow \qquad h = q \wedge c\,, \tag{44}$$

$$[F] = 0\,,\ [dF] = q \wedge f = 0 \qquad \Rightarrow \qquad f = q \wedge a\,. \tag{45}$$

Here c and a are arbitrary 1-forms.

We use the spacetime relation and find for the jumps of the field derivatives

$$h = \kappa(f) = \widetilde{\kappa}(f) + \alpha f\,, \tag{46}$$

with $\widetilde{\kappa} := {}^{(1)}\kappa + {}^{(2)}\kappa$. Accordingly,[5]

$$q \wedge h = q \wedge \widetilde{\kappa}(q \wedge a) = 0\,. \tag{47}$$

This equation is a 3-form with 4 components. We have to solve it with respect to a. As a first step, we have to remove the gauge freedom $a \to a + q\,\varphi$ present in (47). We choose the gauge $\vartheta^{\hat{0}} \stackrel{*}{=} q$. After some heavy algebra, we find (see [20] for details, $a, b, ... = 1, 2, 3$)

$$W^{ab} a_b \stackrel{*}{=} 0\,, \quad \text{with} \quad W^{ab} := \widetilde{\chi}^{\hat{0}a\hat{0}b}\,. \tag{48}$$

These are 3 equations for three a_b's! Nontrivial solutions exist provided

$$\mathcal{W} := \det W^{ab} \stackrel{*}{=} \frac{1}{3!}\hat{\epsilon}_{abc}\hat{\epsilon}_{def} W^{ad} W^{be} W^{cf} \stackrel{*}{=} 0. \tag{49}$$

We can rewrite the latter equation in a manifestly 4-dimensional covariant form ($\hat{\epsilon}_{abc} \equiv \hat{\epsilon}_{\hat{0}abc}$, $e_i{}^{\hat{0}} \stackrel{*}{=} q_i$),

$$\mathcal{W} = \frac{\theta^2}{4!}\,\hat{\epsilon}_{mnpq}\,\hat{\epsilon}_{rstu}\,\widetilde{\chi}^{mnri}\,\widetilde{\chi}^{jpsk}\,\widetilde{\chi}^{lqtu}\,q_i q_j q_k q_l = 0\,,$$

with $\theta := \det(e_i{}^{\alpha})$. The 4-dimensional tensorial transformation behavior is obvious.

[5] Compare the corresponding tensor analytical formula $\partial_\beta \widetilde{\chi}^{\alpha\beta\gamma\delta}\partial_\gamma A_\delta = 0$ (see Post [50], (9.40) for $\chi^{[\alpha\beta\gamma\delta]} = 0$).

We define 4th-order Tamm–Rubilar (TR) tensor density of weight +1,

$$\boxed{\mathcal{G}^{ijkl}(\chi) := \frac{1}{4!}\,\hat{\epsilon}_{mnpq}\,\hat{\epsilon}_{rstu}\,\chi^{mnr(i}\,\chi^{j|ps|k}\,\chi^{l)qtu}\,.} \tag{50}$$

It is totally symmetric $\mathcal{G}^{ijkl}(\chi) = \mathcal{G}^{(ijkl)}(\chi)$. Thus, it has 35 independent components. Because $\chi^{ijkl} = \widetilde{\chi}^{ijkl} + \alpha\,\epsilon^{ijkl}$, the total antisymmetry of ϵ yields $\mathcal{G}(\chi) = \mathcal{G}(\widetilde{\chi})$. An explicit calculation shows that

$$\mathcal{G}^{ijkl}(\chi) = \mathcal{G}^{ijkl}({}^{(1)}\chi) + {}^{(1)}\chi^{m(i|n|j}\,\delta_m^{\ k}\,\delta_n^{\ l)}\,. \tag{51}$$

Summarizing, we find that the wave propagation is governed by the extended Fresnel equation that is generally covariant in 4 dimensions:

$$\boxed{\mathcal{G}^{ijkl}(\widetilde{\chi})\,q_i q_j q_k q_l = 0\,.} \tag{52}$$

The wave covectors q lie on a *quartic Fresnel wave surface*, not exactly what we are observing in vacuum at the present epoch of our universe. Some properties of the TR-tensor, see [53], were discussed recently by Beig [3].

Extended Fresnel Equation Decomposed into Time and Space

Recalling the '6-vector' form of the spacetime relation (34) with the 3×3 constitutive matrices (36) and (37), we can decompose the TR-tensor into time and space pieces: $\mathcal{G}^{0000} =: M$, $\mathcal{G}^{000a} =: \frac{1}{4}M^a$, $\mathcal{G}^{00ab} =: \frac{1}{6}M^{ab}$, $\mathcal{G}^{0abc} =: \frac{1}{4}M^{abc}$, $\mathcal{G}^{abcd} =: M^{abcd}$. Then the Fresnel equation (52) reads

$$q_0^4 \underbrace{M}_{M_0} + q_0^3 \underbrace{q_a\,M^a}_{M_1} + q_0^2 \underbrace{q_a q_b\,M^{ab}}_{M_2} + q_0 \underbrace{q_a q_b q_c\,M^{abc}}_{M_3} + \underbrace{q_a q_b q_c q_d\,M^{abcd}}_{M_4} = 0\,, \tag{53}$$

or

$$M_0\,q_0^4 + M_1\,q_0^3 + M_2\,q_0^2 + M_3\,q_0 + M_4 = 0\,, \tag{54}$$

with

$$M = \det \mathfrak{A}\,, \qquad M^a = -\hat{\epsilon}_{bcd}\left(\mathfrak{A}^{ba}\,\mathfrak{A}^{ce}\,\mathfrak{C}^d_{\ e} + \mathfrak{A}^{ab}\,\mathfrak{A}^{ec}\,\mathfrak{D}_e^{\ d}\right)\,, \tag{55}$$

$$\begin{aligned} M^{ab} &= \frac{1}{2}\,\mathfrak{A}^{(ab)}\left[(\mathfrak{C}^d_{\ d})^2 + (\mathfrak{D}_c^{\ c})^2 - (\mathfrak{C}^c_{\ d} + \mathfrak{D}_d^{\ c})(\mathfrak{C}^d_{\ c} + \mathfrak{D}_c^{\ d})\right] \\ &\quad + (\mathfrak{C}^d_{\ c} + \mathfrak{D}_c^{\ d})(\mathfrak{A}^{c(a}\mathfrak{C}^{b)}_{\ d} + \mathfrak{D}_d^{\ (a}\mathfrak{A}^{b)c}) - \mathfrak{C}^d_{\ d}\mathfrak{A}^{c(a}\mathfrak{C}^{b)}_{\ c} \\ &\quad - \mathfrak{D}_c^{\ (a}\mathfrak{A}^{b)c}\mathfrak{D}_d^{\ d} - \mathfrak{A}^{dc}\mathfrak{C}^{(a}_{\ c}\mathfrak{D}_d^{\ b)} \\ &\quad + \left(\mathfrak{A}^{(ab)}\mathfrak{A}^{dc} - \mathfrak{A}^{d(a}\mathfrak{A}^{b)c}\right)\mathfrak{B}_{dc}\,, \end{aligned} \tag{56}$$

$$\begin{aligned} M^{abc} &= \epsilon^{de(c|}\left[\mathfrak{B}_{df}(\mathfrak{A}^{ab)}\,\mathfrak{D}_e^{\ f} - \mathfrak{D}_e^{\ a}\mathfrak{A}^{b)f}) + \mathfrak{B}_{fd}(\mathfrak{A}^{ab)}\,\mathfrak{C}^f_{\ e} - \mathfrak{A}^{f|a}\mathfrak{C}^{b)}_{\ e}\right) \\ &\quad + \mathfrak{C}^a_{\ f}\,\mathfrak{D}_e^{\ b)}\,\mathfrak{D}_d^{\ f} + \mathfrak{D}_f^{\ a}\,\mathfrak{C}^{b)}_{\ e}\,\mathfrak{C}^f_{\ d}\Big]\,, \end{aligned} \tag{57}$$

$$M^{abcd} = \epsilon^{ef(c}\epsilon^{|gh|d}\,\mathfrak{B}_{hf}\left[\frac{1}{2}\,\mathfrak{A}^{ab)}\,\mathfrak{B}_{ge} - \mathfrak{C}^a_{\ e}\,\mathfrak{D}_g^{\ b)}\right]\,. \tag{58}$$

Fresnel Wave Surfaces

Let us look at some Fresnel wave surfaces in order to get some feeling for the physics involved. Divide (53) by q_0^4 (here q_0 is the frequency of the wave) and introduce the dimensionless variables (c = velocity of light in special relativity)

$$x_a := c\,\frac{q_a}{q_0}\,. \tag{59}$$

Then we have

$$M + x_a \frac{M^a}{c} + x_a x_b \frac{M^{ab}}{c^2} + x_a x_b x_c \frac{M^{abc}}{c^3} + x_a x_b x_c x_d \frac{M^{abcd}}{c^4} = 0\,. \tag{60}$$

We can draw these *quartic* surfaces in the dimensionless variables $x = x_1$, $y = x_2$, $z = x_3$, provided the M's are given. According to (55)–(58), the M's can be expressed in terms of the 3×3 matrices $\mathfrak{A}, \mathfrak{B}, \mathfrak{C}, \mathfrak{D}$. These matrices are specified in (38)–(40) in terms of the permittivity etc.. A comparison with the spacetime relations in the form of (41), (42) is particularly instructive.

Let us start with a simple example. We assume that the permittivity is anisotropic but still diagonal, $\varepsilon^{ab} = \mathrm{diag}(\varepsilon_1, \varepsilon_2, \varepsilon_3)$, whereas the impermeability is trivial $\mu^{-1}_{ab} = \mu_0^{-1}\mathrm{diag}(1,1,1)$. No skewon field is assumed to exist. Whether an axion field is present or not doesn't matter since the axion does not influence the light propagation in the geometrical optics limit. With Mathematica programs written by Tertychniy [59], we can construct for any values of $\varepsilon_1, \varepsilon_2, \varepsilon_3$ the Fresnel wave surface; an example is displayed in Fig. 4.

More complicated cases are trivial permittivity $\varepsilon^{ab} = \varepsilon_0\,\mathrm{diag}(1,1,1)$ and trivial impermeability $\mu^{-1}_{ab} = \mu_0^{-1}\mathrm{diag}(1,1,1)$, but a nontrivial skewon field. We can take a skewon field of electric Faraday type ${\mathcal{S}_3}^0$, for example, see Fig. 5, or of magnetoelectric optical activity type ${\mathcal{S}_1}^2 = {\mathcal{S}_2}^1$, see Fig. 6. In both figures and in the subsequent one $\lambda_0 = \sqrt{\varepsilon_0/\mu_0}$ is the admittance of free space. The characteristic feature of the skewon field is the emergence of specific *holes in the Fresnel surfaces* that correspond to the directions in space along which the wave propagation is damped out completely [45]. This effect is in agreement with our earlier conclusion on the dissipative nature of the skewon field.

Now we can combine anisotropic permittivity with the presence of a skewon field. Then we expect to find some kind of Fig. 4 "enriched" with holes induced by the skewon field. This time we choose a spatially isotropic skewon field with ${\mathcal{S}_1}^1 = {\mathcal{S}_2}^2 = {\mathcal{S}_3}^3 = -\frac{1}{3}\ {\mathcal{S}_0}^0 \neq 0$. The outcome is depicted in Fig. 7. The four holes confirm our expectation.

11 No Birefringence in Vacuum and the Light Cone

The propagation of light in *local* and *linear* premetric vacuum electrodynamics is characterized by the extended Fresnel equation (52) or (54). We can solve the Fresnel equation with respect to the frequency q_0, keeping the 3-covector q_a fixed. With the help of Mathematica, we found the following four solutions [33]:

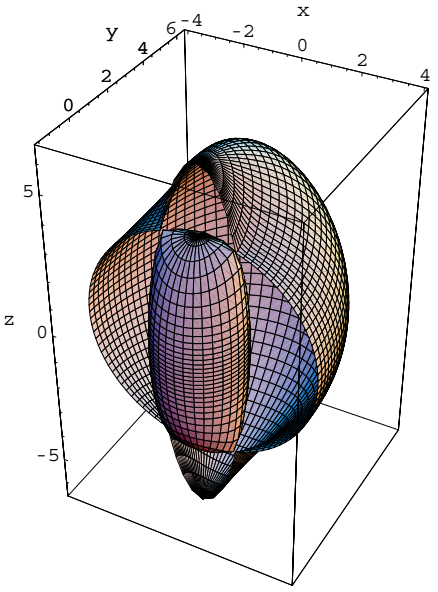


Fig. 4. Fresnel wave surface for anisotropic permittivity $\varepsilon^{ab} = \mathrm{diag}(39.7, 15.4, 2.3)$ and trivial impermeability $\mu_{ab}^{-1} = \mu_0^{-1}\mathrm{diag}(1,1,1)$. The skewon field vanishes. There are two branches, the outer part of the surface is cut into half in order to show the inner branch. We use the dimensionless variables $x := cq_1/q_0$, $y := cq_2/q_0$, $z := cq_3/q_0$

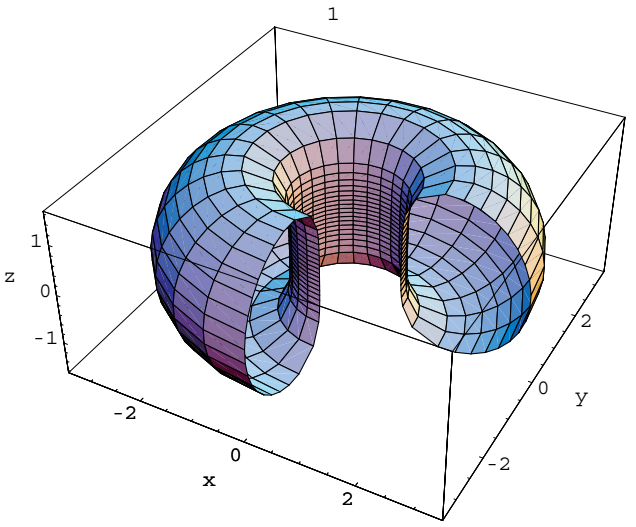


Fig. 5. Fresnel wave surface for trivial permittivity $\varepsilon^{ab} = \varepsilon_0\,\mathrm{diag}(1,1,1)$ and trivial impermeability $\mu_{ab}^{-1} = \mu_0^{-1}\mathrm{diag}(1,1,1)$ with a skewon field of electric Faraday type $\overset{\circ}{\mathcal{S}}{}_3{}^0 = 3.1\lambda_0$ (all other components vanish). The surface has the form of a toroid (depicted with two cuts). We use the dimensionless variables $x := cq_1/q_0$, $y := cq_2/q_0$, $z := cq_3/q_0$

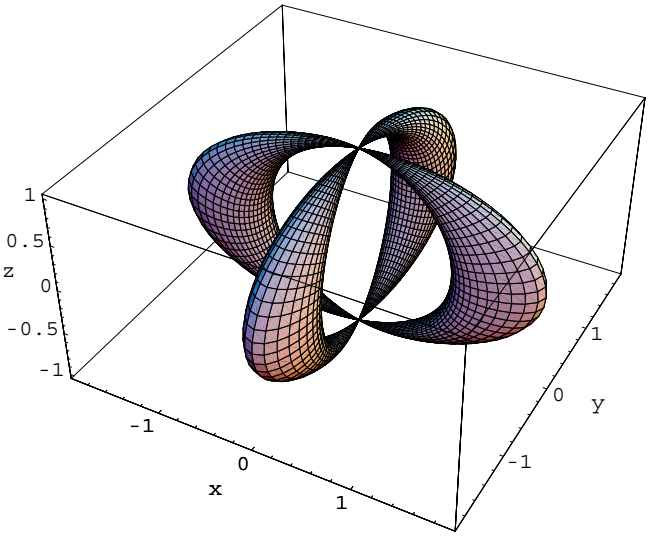


Fig. 6. Fresnel wave surface for trivial permittivity $\varepsilon^{ab} = \varepsilon_0\,\mathrm{diag}(1,1,1)$ and trivial impermeability $\mu^{-1}_{ab} = \mu_0^{-1}\mathrm{diag}(1,1,1)$ with a skewon field of the magneto-electric optical activity type $\not{S}_1{}^2 = \not{S}_2{}^1 = 0.8\,\lambda_0$ (all other components vanish). It has two intersecting toroidal branches. We use the dimensionless variables $x := cq_1/q_0$, $y := cq_2/q_0$, $z := cq_3/q_0$

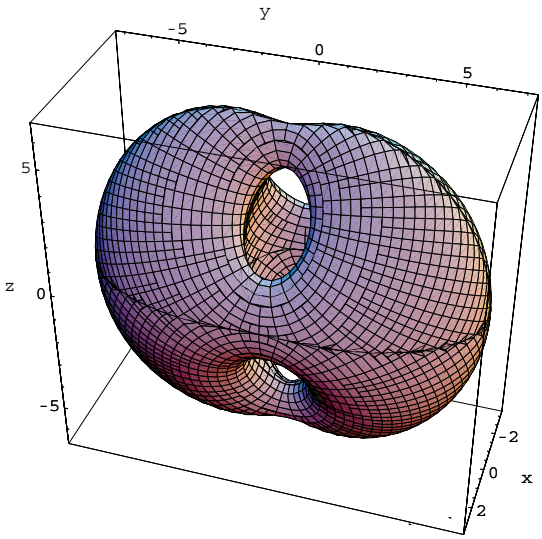


Fig. 7. Fresnel wave surface for anisotropic permittivity $\varepsilon^{ab} = \mathrm{diag}(2.4, 14.8, 54)$ and trivial impermeability $\mu^{-1}_{ab} = \mu_0^{-1}\mathrm{diag}(1,1,1)$ with a spatially isotropic skewon field $\not{S}_1{}^1 = \not{S}_2{}^2 = \not{S}_3{}^3 = -\frac{1}{3}\ \not{S}_0{}^0 = 0.25\,\lambda_0$ (all other components vanish). We use the dimensionless variables $x := cq_1/q_0$, $y := cq_2/q_0$, $z := cq_3/q_0$

$$q^{\uparrow}_{0(\pm)} = \sqrt{\alpha} \pm \sqrt{\beta + \frac{\gamma}{\sqrt{\alpha}}} - \delta\,, \tag{61}$$

$$q^{\downarrow}_{0(\pm)} = -\sqrt{\alpha} \pm \sqrt{\beta - \frac{\gamma}{\sqrt{\alpha}}} - \delta\,. \tag{62}$$

We introduced the abbreviations

$$\alpha := \frac{1}{12M_0}\left(\frac{a}{\left(b+\sqrt{c}\right)^{\frac{1}{3}}} + \left(b+\sqrt{c}\right)^{\frac{1}{3}} - 2M_2\right) + \delta^2\,, \tag{63}$$

$$\beta := \frac{1}{12M_0}\left(-\frac{a}{\left(b+\sqrt{c}\right)^{\frac{1}{3}}} - \left(b+\sqrt{c}\right)^{\frac{1}{3}} - 4M_2\right) + 2\delta^2\,, \tag{64}$$

$$\gamma := \frac{1}{4M_0}\left(2\delta M_2 - M_3\right) - 2\delta^3\,, \qquad \delta := \frac{M_1}{4M_0}\,, \tag{65}$$

with

$$a := 12M_0M_4 - 3M_1M_3 + M_2^2\,, \tag{66}$$

$$b := \frac{27}{2}M_0M_3^2 - 36M_0M_2M_4 - \frac{9}{2}M_1M_2M_3 + \frac{27}{2}M_1^2M_4 + M_2^3\,, \tag{67}$$

$$c := 4\left(b^2 - a^3\right)\,. \tag{68}$$

Vanishing Birefringence

Now, let us demand the absence of birefringence (also called double refraction).[6] In technical terms this means, see the solutions (61), (62), that $\beta = 0$ and $\gamma = 0$. Then we have the degenerate solution

$$q^{\uparrow}_0 = \sqrt{\alpha} - \frac{M_1}{4M_0}\,, \qquad q^{\downarrow}_0 = -\sqrt{\alpha} - \frac{M_1}{4M_0}\,. \tag{69}$$

The condition $\gamma = 0$ yields directly $M_3 = M_1\left(4M_0M_2 - M_1^2\right)/8M_0^2$, and, using this, we find

$$\alpha = \frac{3M_1^2 - 8M_0M_2}{16M_0^2}\,. \tag{70}$$

Thus,

$$q^{\uparrow\downarrow}_0 = \pm\sqrt{\frac{3M_1^2 - 8M_0M_2}{16M_0^2}} - \frac{M_1}{4M_0}\,. \tag{71}$$

Accordingly, the quartic wave surface (54) in this case reduces to

$$[(q_0 - q^{\uparrow}_0)(q_0 - q^{\downarrow}_0)]^2 = 0\,. \tag{72}$$

[6] Similar considerations on vanishing birefringence, for weak gravitational fields, are due to Ni [44]. He was also the first to understand that the axion field doesn't influence light propagation in the geometrical optics limit.

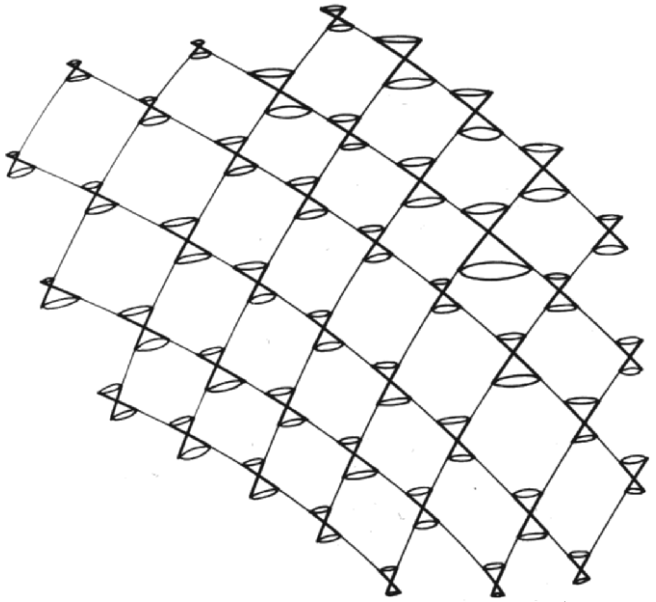


Fig. 8. Null cones fitted together to form a conformal manifold (see Pirani and Schild [49])

Multiplication yields

$$q_0^2 + \frac{1}{2}\,\frac{M_1}{M_0}\,q_0 + \frac{1}{2}\,\frac{M_2}{M_0} - \frac{1}{8}\left(\frac{M_1}{M_0}\right)^2 = 0\,. \tag{73}$$

If we substitute M_0, M_1, M_2 as defined in (53), we have explicitly $(i, j = 0, 1, 2, 3)$

$$g^{ij}q_i q_j := q_0^2 + \frac{1}{2}\,\frac{M^a}{M}\,q_0 q_a + \frac{1}{8}\left(4\,\frac{M^{ab}}{M} - \frac{M^a M^b}{M^2}\right) q_a q_b = 0\,. \tag{74}$$

This equation is quadratic in the 4-dimensional wave covector q_i. Therefore we recover the conventional *light cone* of general relativity at each point of spacetime, see Fig. 8. Thereby the *causal structure* of spacetime is determined. Thus, up to a scalar factor, we derived the Riemannian metric of general relativity.

Moreover, as we have shown [20,25], we find the correct *Lorentzian signature*. The Lorentzian (also known as Minkowskian) signature can be traced back to the Lenz rule, which determines the sign of the $\dot{B}$ term in the induction law.[7] And this sign is different from the one in the corresponding $\dot{\mathcal{D}}$ term in the Oersted-Ampère-Maxwell law. In other words, the Lorentz signature is encoded

[7] Usually it is argued that the signature should be derived from quantum field theoretical principles; for a corresponding model, see, e.g., Froggatt & Nielsen [15]. Needless to say that it is our view that classical premetric electrodynamics together with the Lenz rule and a local and linear spacetime relation is all what is really needed.

in the decomposition formulas (9) and (14). Neither is the minus sign in (9) a convention nor the plus sign of the $E \wedge d\sigma$ term in (14). Since the Lenz rule is related to the positivity of the electromagnetic energy, the same is true for the Lorentzian signature. This derivation of the signature of the metric of spacetime from electrodynamics provides new insight into the structures underlying special as well as general relativity.

12 Dilaton, Metric, Axion

At first the skewon and the axion emerged at the premetric level in our theory and only subsequently the metric. Consequently, the axion and the skewon should be regarded as more fundamental fields (if they exist) than the metric. In the meantime, we phased out the skewon field since we insisted, in Sect. 11, on vanishing birefringence in vacuum.

As to the metric, we recognize that multiplication of the metric by an arbitrary function $\tilde{\lambda}(x)$ was left open in the derivation of the last section, see (74):

$$\tilde{\lambda}(x)\, g^{ij}(x)\, q_i q_j = 0\,. \tag{75}$$

Thus, only the conformally invariant part of the metric is determined. In other words, we have actually constructed the *conformal* (or the light cone) *structure* on the spacetime manifold, see, e.g., Weyl [63, 64], Schouten [55], and Pirani & Schild [49].

It is known from *special relativity* that the light cone (with Lorentzian signature) is invariant under the 15-parameter conformal group, see Barut & Rączka [2] and Blagojević [4]. The latter, in Minkowskian coordinates x^i, is generated by the following four sets of spacetime transformations:

$$\text{Translations} \quad (4 \text{ param.}) \qquad x^i \to \tilde{x}^i = x^i + a^i\,, \tag{76}$$

$$\text{Lorentz transf.} \quad (6 \text{ param.}) \qquad x^i \to \tilde{x}^i = \Lambda^i{}_j\, x^j\,, \tag{77}$$

$$\text{dila(ta)tion} \quad (1 \text{ param.}) \qquad x^i \to \tilde{x}^i = \rho\, x^i\,, \tag{78}$$

$$\text{prop. conf. transf.} \quad (4 \text{ param.}) \qquad x^i \to \tilde{x}^i = \frac{x^i + \kappa^i\, x^2}{1 + 2\kappa_j\, x^j + \kappa_j \kappa^j\, x^2}. \tag{79}$$

Here $a^i, \Lambda^i{}_j, \rho, \kappa^i$ are the 15 constant parameters, and $x^2 := g_{ij} x^i x^j$. The Poincaré subgroup (76), (77) (for a modern presentation of it, see Giulini [17]) leaves the spacetime interval $ds^2 = g_{ij} dx^i dx^j$ invariant, whereas the dilatations (78) and the proper conformal transformations (79) change the spacetime interval by a scaling factor $ds^2 \to \rho^2 ds^2$ and $ds^2 \to \sigma^2 ds^2$, respectively (with $\sigma^{-1} := 1 + 2\kappa_j\, x^j + \kappa_j \kappa^j\, x^2$). In all cases the light cone $ds^2 = 0$ is left invariant. The Weyl subgroup, which is generated by the transformations (76)–(78), and its corresponding Noether currents were discussed by, e.g., Kopczyński et al. [30].

For massless particles, instead of the Poincaré group, the conformal or the Weyl group come under consideration, since massless particles move on the light

cone. Even though the light cone stays invariant under all transformations (76)–(79), two reference frames that are linked to each other by a proper conformal transformation don't stay inertial frames since their relative velocity is not constant. If one wants to uphold the *inertial* character of the reference frames, one has to turn to the Weyl transformation, that is, one has to specialize to $\kappa^i = 0$.

The conformal group in Minkowski space illustrates the importance of the light cone structure on a *flat* manifold. This is suggestive for the light cone on an arbitrarily curved manifold, even though there is no direct relation between (76)–(79) and the light cone structure we derived in the last section.

The light cone metric g^{ij} introduces the Hodge star * operator. We then can straightforwardly verify that the principal part of the spacetime relation is determined as $H \sim {}^\star F$, where the coefficient of proportionality can be an arbitrary scalar function $\lambda(x)$ of the spacetime coordinates. This function is naturally identified with the *dilaton field*, see Brans [6] and Fujii & Maeda [16]. Introducing the (Levi-Civita) dual of the excitation, $\check{H}^{ij} := \frac{1}{2}\,\epsilon^{ijkl}\,H_{kl}$, we can then finally rewrite the spacetime relation for vanishing birefringence in vacuum as

$$\check{H}^{ij} = [\ \underbrace{\lambda(x)}_{\text{dilaton}}\ \sqrt{-g}\,g^{ik}(x)\,g^{jl}(x) + \underbrace{\alpha(x)}_{\text{axion}}\ \epsilon^{ijkl}\ ]\,F_{kl}\,, \tag{80}$$

that is, we are left with the constitutive fields dilaton λ, metric g^{ij}, and axion α. The combination $\sqrt{-g}\,g^{i[k}(x)\,g^{l]j}(x)$ is conformally invariant, in complete agreement with the above analysis.

13 Setting the Scale

The conformal structure of spacetime is laid down in (74). Hence only 9 of the 10 independent components of the pseudo-Riemannian metric g_{ij} are specified. We need, in addition to the conformal structure, a *volume measure* for arriving at a unique Riemannian metric. This can be achieved by postulating a time or length standard.

In exterior calculus, (80) reads

$$H = \lambda(x)\,{}^\star F + \alpha(x)\,F\,. \tag{81}$$

The axion has not been found so far, so we can put $\alpha = 0$. Moreover, under normal cicumstances, the dilaton seems to be a constant field and thereby sets a certain scale, i.e., $\lambda(x) = \lambda_0$, where λ_0 is the admittance of free space[8] the value of which is, in SI-units, $1/(377\ \Omega)$. (The exact implementation of this assumption will have to be worked out in future.) Accordingly, we are left with the spacetime relation of conventional Maxwell–Lorentz electrodynamics

$$\boxed{H = \lambda_0\,{}^\star F} \qquad \text{or} \qquad \check{H}^{ij} = \lambda_0\sqrt{-g}\,g^{ik}(x)\,g^{jl}(x)\,F_{kl} = \lambda_0\sqrt{-g}\,F^{ij}\,. \tag{82}$$

[8] Our electrodynamical formalism is independent of the chosen system of units, as we discussed elsewhere [22].

14 Discussion

Weyl [63, 64], in 1921, proved a theorem that the projective and the conformal structures of a metrical space determine its metric uniquely. As a consequence Weyl [63] argued that *...the metric of the world can be determined merely by observing the "natural" motion of material particles and the propagation of action, in particular that of light; measuring rods and clocks are not required for that.* Here we find the two elementary notions for the determination of the metric: The paths of a freely falling point particles, yielding the projective structure, and light rays, yielding the conformal structure of spacetime. Later, in 1966, Pirani and Schild [49], amongst others, deepened the insight into the conformal structure and the Weyl tensor.

In 1972, on the basis of Weyl's two primitive elements, Ehlers, Pirani, and Schild (EPS) [12] proposed an axiomatic framework in which Weyl's concepts of free particles and of light rays were taken as elementary notions that are linked to each other by plausible axioms. Requiring compatibility between the emerging projective and conformal structures, they ended up with a Weyl spacetime[9](Riemannian metric with an additional Weyl covector). They set a scale [as we did in the last section, too] and arrived at the pseudo-Riemannian metric of general relativity. In this sense, EPS were able to reconstruct the metric of general relativity.

Subsequently, many authors improved and discussed the EPS-axiomatics. Access to the corresponding literature can be found via the book of Majer and Schmidt [40] or the work of Perlick [47, 48] and Lämmerzahl [31], e.g.. For a general review one should compare Schelb [54] and for a new axiomatic scheme Schröter [56].

As stated, the point particles and light rays were primary elements that were assumed to exist and no link to mechanics nor to electrodynamics was specified. The particle concept within the EPS-axiomatics lost credibility when during the emergence of gauge theories of gravity (which started in 1956 with Utiyama [61] even before the EPS-framework had been set up in 1972) the first quantized wave function Ψ for matter entered the scene as an elementary and "irreducible" concept in gravity theory. When *neutron* interference in an external gravitational field was discovered experimentally in 1975 by Collella, Overhauser, and Werner (COW) [7], see also [52], Sect. 7, it was clear that the point particle concept in the EPS-framework became untenable from a physical point of view. For completeness let us mention some more recent experiments on matter waves in the gravitational field or in a noninertial frame:

- The Werner, Staudenmann, and Colella experiment [62] in 1979 on the phase shift of neutron waves induced by the rotation of the Earth (Sagnac-type effect),

[9] Time measurement in Weyl spacetime were discussed by Perlick [47] and by Teyssandier & Tucker [60].

- the Bonse & Wroblewski experiment [5] in 1984 on neutron interferometry in a noninertial frame (verifying, together with the COW experiment, the equivalence principle for neutron waves),
- the Kasevich & Chu interferometric experiment [27] in 1991 with laser-cooled wave packets of sodium *atoms* in the gravitational field,
- the Mewes et al. experiment [41] in 1997 with interfering freely falling Bose-Einstein condensed sodium atoms, see Ketterle [28], Fig. 14 and the corresponding text,
- the Nesvizhevsky et al. experiment [42,43] in 2002 on the quantum states of neutrons in the Earth's gravitational field, and
- the Fray, Hänsch, et al. [14] experiment in 2004 with a matter wave interferometer based on the diffraction of atoms from effective absorption gratings of light. This interferometer was used for two stable isotopes of the rubidium atom in the gravitational field of the Earth. Thereby the equivalence principle was tested successfully on the atomic level.

Clearly, without the Schrödinger equation in an external gravitational field or in a noninertial frame all these experiments cannot be described.[10] Still, in most textbooks on gravity, these experiments are not even mentioned!

In the 1980's, as a reaction to the COW-experiment, Audretsch and Lämmerzahl, for a review see [1], started to develop an axiomatic scheme for spacetime in which the point particle was substituted by a matter wave function and the light ray be a wave equation for electromagnetic disturbances. In this way, they could also include projective structures with an asymmetric connection (i.e., with torsion), which was excluded in the EPS approach a priori.

Turning to the conformal structure, which is in the center of our interest here, Lämmerzahl et al. [32], see also [18,51], reconsidered the Audretsch-Lämmerzahl scheme and derived the inhomogeneous Maxwell equation from the following requirements: a well-posed Cauchy problem, the superposition principle, a finite propagation speed, and the absence of birefringence in vacuum. The homogeneous Maxwell equation they got by a suitable definition of the electromagnetic field strength. With a geometric optics approximation, compare our Sect. 10, they recover the light ray in lowest order. And this is the message of this type of axiomatics: Within the axiomatic system of Audretsch and Lämmerzahl et al., the light ray, which is elementary in the EPS-approach, can be derived from reasonable axioms about the propagation of electromagnetic disturbances. As with the substitution of the mass point by a matter wave, this inquiry into the physical nature of the light ray and the corresponding reshaping of the EPS-scheme seems to lead to a better understanding of the metric of spacetime. And this is exactly where our framework fits in: We also build up the Maxwell equations in an axiomatic way and are even led to the signature of the metric, an achievement that needs still to be evaluated in all details.

[10] A systematic procedure of deriving the COW result by applying the equivalence principle to the Dirac equation can be found in [19].

15 Summary

Let us then summarize our findings: We outlined our axiomatic approach to electrodynamics and to the derivation of the light cone. In particular, with the help of a local and linear spacetime relation,

- we found the skewon field $\not{S}_i{}^j$ (15 components) and the axion field α (1 component),
- we found a *quartic* Fresnel wave surface for light propagation.
- In the case of vanishing *birefringence,* the Fresnel wave surface degenerates and we recovered the light cone (determining 9 components of the metric tensor) and, together with it, the conformal and causal structure of spacetime and the Hodge star * operator.
- If additionally the dilaton λ (1 component) is put to a constant, namely to the admittance of free space λ_0 [$1/(377\ \Omega)$ in SI-units], and the axion α removed, we recover the conventional Maxwell–Lorentz spacetime relation $H = \lambda_0\, {}^\star F$.

Thus, in our framework, the conformal part of the *metric* emerges from the local and linear spacetime relation as an *electromagnetic construct.* In this sense, the light cone is a derived concept.

Acknowledgments

One of us is grateful to Claus Lämmerzahl and Jürgen Ehlers for the invitation to the Potsdam Seminar. Many thanks go to Sergey Tertychniy (Moscow) who provided the Mathematica programs for the drawing of the Fresnel surfaces. Financial support from the DFG, Bonn (HE-528/20-1) and from INTAS, Brussels, is gratefully acknowledged.

References

1. J. Audretsch and C. Lämmerzahl, *A new constructive axiomatic scheme for the geometry of space-time* In [40] pp. 21–39 (1994).
2. A.O. Barut and R. Rączka, *Theory of Group Representations and Applications* (PWN – Polish Scientific Publishers, Warsaw, 1977).
3. R. Beig, *Concepts of Hyperbolicity and Relativistic Continuum Mechanics,* arXiv.org/gr-qc/0411092.
4. M. Blagojević, *Gravitation and Gauge Symmetries* (IOP Publishing, Bristol, 2002).
5. U. Bonse and T. Wroblewski, *Dynamical diffraction effects in noninertial neutron interferometry, Phys. Rev.* **D30** (1984) 1214–1217.
6. C.H. Brans, *The roots of scalar-tensor theory: An approximate history,* arXiv.org/gr-qc/0506063.
7. R. Colella, A.W. Overhauser, and S.A. Werner *Observation of Gravitationally Induced Quantum Interference*, *Phys. Rev. Lett.* **34** (1975) 1472–1474.

8. D.H. Delphenich, *On the axioms of topological electromagnetism, Ann. Phys. (Leipzig)* **14** (2005) 347–377; updated version of arXiv.org/hep-th/0311256.
9. D.H. Delphenich, *Symmetries and pre-metric electromagnetism, Ann. Phys. (Leipzig)* **14** (2005) issue 11 or 12, to be published.
10. I.E. Dzyaloshinskii, *On the magneto-electrical effect in antiferromagnets, J. Exptl. Theoret. Phys. (USSR)* **37** (1959) 881–882 [English transl.: *Sov. Phys. JETP* **10** (1960) 628–629].
11. D.G.B. Edelen, *A metric free electrodynamics with electric and magnetic charges, Ann. Phys. (NY)* **112** (1978) 366–400.
12. J. Ehlers, F.A.E. Pirani, and A. Schild, *The geometry of free fall and light propagation,* in: *General Relativity, papers in honour of J.L. Synge*, L. O'Raifeartaigh, ed. (Clarendon Press, Oxford, 1972), pp. 63–84.
13. A. Einstein, *Zur Elektrodynamik bewegter Körper, Ann. Phys. (Leipzig)* **17** (1905) 891–921. English translation in [39].
14. S. Fray, C. Alvarez Diez, T. W. Hänsch and M. Weitz, *Atomic interferometer with amplitude gratings of light and its applications to atom based tests of the equivalence principle, Phys. Rev. Lett.* **93** (2004) 240404 (4 pages); arXiv.org/physics/0411052.
15. C.D. Froggatt and H.B. Nielsen, *Derivation of Poincaré invariance from general quantum field theory, Ann. Phys. (Leipzig)* **14** (2005) 115–147 [Special issue commemorating Albert Einstein].
16. Y. Fujii and K.-I. Maeda, *The Scalar-Tensor Theory of Gravitation* (Cambridge University Press, Cambridge, 2003).
17. D. Giulini, *The Poincaré group: Algebraic, representation-theoretic, and geometric aspects,* these Proceedings (2005).
18. M. Haugan and C. Lämmerzahl, *On the experimental foundations of the Maxwell equations*, *Ann. Phys. (Leipzig)* **9** (2000) Special Issue, SI-119–SI-124.
19. F.W. Hehl and W.-T. Ni: *Inertial effects of a Dirac particle. Phys. Rev.* **D42** (1990) 2045–2048.
20. F.W. Hehl and Yu.N. Obukhov, *Foundations of Classical Electrodynamics — Charge, Flux, and Metric* (Birkhäuser, Boston, MA, 2003).
21. F.W. Hehl and Yu.N. Obukhov, *Electric/magnetic reciprocity in premetric electrodynamics with and without magnetic charge, and the complex electromagnetic field, Phys. Lett.* **A323** (2004) 169–175; arXiv.org/physics/0401083.
22. F.W. Hehl and Yu. N. Obukhov, *Dimensions and units in electrodynamics, Gen. Rel. Grav.* **37** (2005) 733–749; arXiv.org/physics/0407022.
23. F.W. Hehl and Yu.N. Obukhov, *Linear media in classical electrodynamics and the Post constraint, Phys. Lett.* **A334** (2005) 249–259; arXiv.org/physics/ 0411038.
24. Y. Itin, *Caroll-Field-Jackiw electrodynamics in the pre-metric framework, Phys. Rev.* **D70** (2004) 025012 (6 pages); arXiv/org/hep-th/0403023.
25. Y. Itin and F.W. Hehl, *Is the Lorentz signature of the metric of spacetime electromagnetic in origin? Ann. Phys. (NY)* **312** (2004) 60–83; arXiv.org/gr-qc/0401016.
26. G. Kaiser, *Energy-momentum conservation in pre-metric electrodynamics with magnetic charges*, *J. Phys.* **A37** (2004) 7163–7168.
27. M. Kasevich and S. Chu, *Atomic interferometry using stimulated Raman transitions, Phys. Rev. Lett.* **67** (1991) 181–184.
28. W. Ketterle, *When atoms behave as waves: Bose-Einstein condensation and the atom laser* (Noble lecture 2001), `http://nobelprize.org/physics/laureates/2001/ketterle-lecture.pdf` .
29. R.M. Kiehn, *Plasmas and Non Equilibrium Electrodynamics* 2005 (312 pages), see `http://www22.pair.com/csdc/download/plasmas85h.pdf` .

30. W. Kopczyński, J.D. McCrea and F.W. Hehl, *The Weyl group and its current. Phys. Lett.* **A128** (1988) 313–317.
31. C. Lämmerzahl, *A characterisation of the Weylian structure of space-time by means of low velocity tests, Gen. Rel. Grav.* **33** (2001) 815–831; arXiv.org/gr-qc/0103047.
32. C. Lämmerzahl, A. Camacho, and A. Macías, *Reasons for the electromagnetic field to obey Maxwell's equations,* submitted to *J. Math. Phys.* (2005).
33. C. Lämmerzahl and F.W. Hehl, *Riemannian light cone from vanishing birefringence in premetric vacuum electrodynamics*, *Phys. Rev.* **D70** (2004) 105022 (7 pages); arXiv.org/gr-qc/0409072.
34. C. Lämmerzahl, A. Macías, and H. Müller, *Lorentz invariance violation and charge (non)conservation: A general theoretical frame for extensions of the Maxwell equations, Phys. Rev.* **D71** (2005) 025007 (15 pages).
35. I.V. Lindell, *Differential Forms in Electromagnetics* (IEEE Press, Piscataway, NJ, and Wiley-Interscience, 2004).
36. L.V. Lindell, *The class of bi-anisotropic IB-media*, *J. Electromag. Waves Appl.,* to be published (2005).
37. I.V. Lindell and A.H. Sihvola, *Perfect electromagnetic conductor, J. Electromag. Waves Appl.* **19** (2005) 861–869; arXiv.org/physics/0503232.
38. I.V. Lindell, A.H. Sihvola, S.A. Tretyakov, A.J. Viitanen, *Electromagnetic Waves in Chiral and Bi-Isotropic Media* (Artech House, Boston, MA, 1994).
39. H.A. Lorentz, A. Einstein, H. Minkowski, and H. Weyl, *The Principle of Relativity.* A collection of original memoirs. Translation from the German (Dover, New York, 1952).
40. U. Majer and H.-J.Schmidt (eds.), *Semantical Aspects of Spacetime Theories* (BI-Wissenschaftsverlag, Mannheim 1994).
41. M.-O. Mewes et al., *Output Coupler for Bose-Einstein Condensed Atoms, Phys. Rev. Lett.* **78** (1997) 582–585.
42. V.V. Nesvizhevsky et al., *Quantum states of neutrons in the Earth's gravitational field, Nature* **415** (2002) 297–299.
43. V.V. Nesvizhevsky et al., *Measurement of quantum states of neutrons in the Earth's gravitational field, Phys. Rev.* **D67** (2003) 102002 (9 pages); arXiv.org/hep-ph/0306198.
44. W.-T. Ni, *Equivalence principles and precision experiments.* In *Precision Measurement and Fundamental Constants II,* B.N. Taylor, W.D. Phillips, eds. Nat. Bur. Stand. (US) Spec. Publ. 617 (US Government Printing Office, Washington, DC, 1984) pp. 647–651.
45. Yu.N. Obukhov and F.W. Hehl, *Possible skewon effects on light propagation, Phys. Rev.* **D70** (2004) 125015 (14 pages); arXiv.org/physics/0409155.
46. Yu.N. Obukhov and F.W. Hehl, *Measuring a piecewise constant axion field in classical electrodynamics, Phys. Lett.* **A341** (2005) 357–365; arXiv.org/physics/0504172.
47. V. Perlick, *Observer fields in Weylian spacetime models, Class. Quantum Grav.* **8** (1991) 1369–1385.
48. V. Perlick, *Ray optics, Fermat's principle, and applications to general relativity,* Lecture Notes in Physics (Springer) **m61** (2000) 220 pages.
49. F.A.E. Pirani and A. Schild, *Conformal geometry and the interpretation of the Weyl tensor*, in: *Perspectives in Geometry and Relativity.* Essays in honor of V. Hlavatý. B. Hoffmann, editor (Indiana University Press, Bloomington, 1966) pp. 291–309.

50. E.J. Post, *Formal Structure of Electromagnetics – General Covariance and Electromagnetics* (North Holland, Amsterdam, 1962, and Dover, Mineola, New York, 1997).
51. R.A. Puntigam, C. Lämmerzahl and F.W. Hehl, *Maxwell's theory on a post-Riemannian spacetime and the equivalence principle, Class. Quant. Grav.* **14** (1997) 1347–1356; arXiv.org/gr-qc/9607023.
52. H. Rauch and S.A. Werner, *Neutron Interferometry,* Lessons in experimental quantum mechanics (Clarendon Press, Oxford, 2000).
53. G.F. Rubilar, *Linear pre-metric electrodynamics and deduction of the lightcone*, *Thesis* (University of Cologne, June 2002); see *Ann. Phys. (Leipzig)* **11** (2002) 717–782.
54. U. Schelb, *Zur physikalischen Begründung der Raum-Zeit-Geometrie,* Habilitation thesis (Univ. Paderborn, 1997).
55. J.A. Schouten, *Ricci-Calculus*, 2nd ed. (Springer, Berlin, 1954).
56. J. Schröter, *A new formulation of general relativity.* Part I. Pre-radar charts as generating functions for metric and velocity. Part II. Pre-radar charts as generating functions in arbitrary space-times. Part III. GTR as scalar field theory (altogether 58 pages). Preprint, Univ. Paderborn (July 2005).
57. B.D.H. Tellegen, *The gyrator, a new electric network element, Philips Res. Rep.* **3** (1948) 81–101.
58. B.D.H. Tellegen, *The gyrator, an electric network element,* in: *Philips Technical Review* **18** (1956/57) 120–124. Reprinted in H.B.G. Casimir and S. Gradstein (eds.) *An Anthology of Philips Research.* (Philips' Gloeilampenfabrieken, Eindhoven, 1966) pp. 186–190.
59. S.I. Tertychniy [National Research Institute for Physical, Technical, and Radio-Technical Measurements (VNIIFTRI), 141570 Mendeleevo, Russia. Email: `bpt97@mendeleevo.ru`] provided the Mathematica programs for constructing the figures of the quartic Fresnel surface.
60. P. Teyssandier and R. W. Tucker, *Gravity, Gauges and Clocks, Class. Quant. Grav.* **13** (1996) 145–152.
61. R. Utiyama, *Invariant theoretical interpretation of interaction, Phys. Rev.* **101** (1956) 1597–1607.
62. S.A. Werner, J.-L. Staudenmann, and R. Colella, *Effect of Earth's rotation on the quantum mechanical phase of the neutron, Phys. Rev. Lett.* **42** (1979) 1103–1106.
63. H. Weyl, *Zur Infinitesimalgeometrie: Einordnung der projektiven und konformen Auffassung, Nachr. Königl. Gesellschaft Wiss. Göttingen, Math.-Phys. Klasse,* pp. 99–112 (1921); also in K. Chandrasekharan (ed.), *Hermann Weyl, Gesammelte Abhandlungen* Vol.II, 195–207 (Springer, Berlin, 1968).
64. H. Weyl, *Raum, Zeit, Materie*, Vorlesungen über Allgemeine Relativitätstheorie, reprint of the 5th ed. of 1923 (Wissenschaftliche Buchges., Darmstadt, 1961). Engl. translation of the 4th ed.: *Space–Time–Matter* (Dover Publ., New York, 1952).
65. F. Wilczek, *Two applications of axion electrodynamics*, *Phys. Rev. Lett.* **58** (1987) 1799–1802.

Part III

Violations of Lorentz Invariance?

Overview of the Standard Model Extension: Implications and Phenomenology of Lorentz Violation

R. Bluhm

Colby College, Waterville, ME 04901, USA
rtbluhm@colby.edu

Abstract. The Standard Model Extension (SME) provides the most general observer-independent field theoretical framework for investigations of Lorentz violation. The SME lagrangian by definition contains all Lorentz-violating interaction terms that can be written as observer scalars and that involve particle fields in the Standard Model and gravitational fields in a generalized theory of gravity. This includes all possible terms that could arise from a process of spontaneous Lorentz violation in the context of a more fundamental theory, as well as terms that explicitly break Lorentz symmetry. An overview of the SME is presented, including its motivations and construction. Some of the theoretical issues arising in the case of spontaneous Lorentz violation are discussed, including the question of what happens to the Nambu-Goldstone modes when Lorentz symmetry is spontaneously violated and whether a Higgs mechanism can occur. A minimal version of the SME in flat Minkowski spacetime that maintains gauge invariance and power-counting renormalizability is used to search for leading-order signals of Lorentz violation. Recent Lorentz tests in QED systems are examined, including experiments with photons, particle and atomic experiments, proposed experiments in space, and experiments with a spin-polarized torsion pendulum.

1 Introduction

It has been 100 years since Einstein published his first papers on special relativity [1]. This theory is based on the principle of Lorentz invariance, that the laws of physics and the speed of light are the same in all inertial frames. A few years after Einstein's initial work, Minkowski showed that a new spacetime geometry emerges from special relativity. In this context, Lorentz symmetry is an exact spacetime symmetry that maintains the form of the Minkowski metric in different Cartesian-coordinate frames.

In the years 1907–1915, Einstein developed the general theory of relativity as a new theory of gravity. In general relativity, spacetime is described in terms of a metric that is a solution of Einstein's equations. The geometry is Riemannian, and the physics is invariant under general coordinate transformations. Lorentz symmetry, on the other hand, becomes a local symmetry. At each point on the

R. Bluhm: *Overview of the Standard Model Extension: Implications and Phenomenology of Lorentz Violation*, Lect. Notes Phys. **702**, 191–226 (2006)
DOI 10.1007/3-540-34523-X_8 © Springer-Verlag Berlin Heidelberg 2006

spacetime manifold, local coordinate frames can be found in which the metric becomes the Minkowski metric. However, the choice of the local frame is not unique, and local Lorentz transformations provide the link between physically equivalent local frames.

The Standard Model (SM) of particle physics is a fully relativistic theory. The SM in Minkowski spacetime is invariant under global Lorentz transformations, whereas in a Riemannian spacetime the particle interactions must remain invariant under both general coordinate transformations and local Lorentz transformations. Particle fields are also invariant under gauge transformations. Exact symmetry under local gauge transformations leads to the existence of massless gauge fields, such as the photon. However, spontaneous breaking of local gauge symmetry in the electroweak theory involves the Higgs mechanism, in which the gauge fields can acquire a mass.

Classical gravitational interactions can be described in a form analogous to gauge theory by using a vierbein formalism [2]. This also permits a straightforward treatment of fermions in curved spacetimes. Covariant derivatives of tensors in the local Lorentz frame involve introducing the spin connection. In a Riemann spacetime with zero torsion, the spin connection is not an independendent field, but rather is a prescribed function of the vierbein and its derivatives. However, a natural generalization is to treat the spin connection components as independent degrees of freedom. The resulting geometry is a Riemann-Cartan spacetime, which has nonvanishing torsion [3]. In a Riemann-Cartan spacetime, the associated field strengths for the vierbein and spin connection are the curvature and torsion tensors. The usual Riemann spacetime of general relativity is recovered in the zero-torsion limit. Similarly, if the curvature tensor vanishes, the spacetime reduces to Minkowski spacetime.

The combination of the SM and Einstein's classical gravitational theory provides a highly successful description of nature. However, since Einstein's theory is not a quantum theory, it is expected that it will ultimately be superseded by a more fundamental theory that will hold at the quantum level. Candidate quantum gravity theories include string theory and loop quantum gravity. The appropriate scale where gravity and quantum physics are expected to meet up is the Planck scale, $m_P \simeq 10^{19}$ GeV.

Finding experimental confirmation of a quantum theory of gravity by doing experiments at the Planck scale, however, is not practical. Instead, an alternative approach can be adopted in which one looks for small Planck-suppressed effects of new physics that might be observable in high-precision experiments. For this idea to hold, any new effect would have to be one that cannot be mimicked by known conventional processes in the SM or conventional gravity theory. One possible signal fulfilling this requirement is to look for Planck-suppressed signatures of Lorentz violation in high-precision experiments.

Detection of such a violation of relativity theory would clearly be a dramatic indication of new physics, presumably coming from the Planck scale. This idea is not merely speculative because it has been shown that mechanisms in both string theory [4, 5] and quantum gravity [6] can lead to violations of Lorentz

symmetry. However, these theories are not yet sufficiently developed in such a way that allows testable predictions to be made at a definite (quantifiable) scale at low energies.

Nonetheless, progress can still be made using effective field theory. To be realistic, an effective field theory would have to contain both the SM and general relativity together with any higher-order couplings between them. It should also maintain coordinate (or observer) independence. In full generality, the gravity sector could include additional fields such as torsion that are not a part of Einstein's general relativity. This would permit more general geometries, including a Riemann-Cartan spacetime.

The general effective field theory of this type incorporating arbitrary observer-independent Lorentz violation is called the Standard-Model Extension (SME) [7–9]. The SME lagrangian by definition contains all observer-scalar terms consisting of products of SM and general gravitational fields with each other as well as with additional couplings that introduce violations of Lorentz symmetry. In principle, there are an infinity of terms in the SME, including nonrenormalizable terms of arbitrarily high dimension.

To investigate low-energy experiments, where the leading-order signals of Lorentz violation are of primary interest, it is often advantageous to work with a subset of the full SME, which includes only a finite number of terms. One subset in particular, referred to as the minimal SME, restricts the theory to power-counting renormalizable and gauge-invariant terms. In recent years, the Lorentz-violating coefficients in the minimal SME have been adopted by experimentalists as the standard for reporting bounds on Lorentz violation. Since many of the low-energy experiments involve only electromagnetic interactions between charged particles and photons, it often suffices to define a minimal QED sector of the SME.

This paper is intended as an overview in the context of the SME of some recent theoretical and phenomenological investigations of Lorentz violation. The motivations for the development of the SME are presented first. An outline of how the theory is constructed is then given. This is followed by a discussion of some theoretical issues that come up when Lorentz violation is due to a process of spontaneous symmetry breaking. In particular, the fate of the Nambu-Goldstone modes is examined along with the question of whether a Higgs mechanism can occur [10]. For simplicity, this discussion is carried out in the context of a vector model known as a bumblebee model [9,11]. The role of the geometry (Minkowski, Riemann, or Riemann-Cartan) is examined as well. To investigate phenomenology, the minimal SME is constructed and used to examine a wide range of experiments assuming a flat Minkowski background. In this paper, the focus is on high-precision tests in QED systems. A number of recent experiments in atomic and particle systems are examined, and the status of their attainable sensitivities to Lorentz violation is reviewed.

The SME is the result of a large on-going collaboration by a group of theorists and experimentalists most of whom have in common that they have at some point collaborated with Alan Kostelecky at Indiana University. An exhaustive review

covering all of this collective work, which spans topics in field theory, gravity, astrophysics, cosmology, as well as particle, nuclear, and atomic physics, is not possible here. Instead, this review focuses mostly on selective recent topics that are of interest to the author. It is also not possible here to give a complete list of references on all of the work looking at possible violations or tests of relativity. For that, other recent reviews and proceedings collections should be consulted as well. See, for example, [12–14].

2 Motivations

Historically, interest in Lorentz violation increased dramatically after it was discovered by Kostelecky and Samuel in the late 1980s that mechanisms can occur in string field theory that could cause spontaneous breaking of Lorentz symmetry [4]. It is this idea that ultimately led to the development of the SME, which in turn has stimulated a variety of experimental searches for relativity violations.

Spontaneous Lorentz violation can occur when a string field theory has a nonperturbative vacuum that can lead to tensor-valued fields acquiring nonzero vacuum expectation values (vevs), $\langle T \rangle \neq 0$. As a result of this, the low-energy effective theory contains an unlimited number of terms of the form

$$\mathcal{L} \sim \frac{\lambda}{m_P^k} \langle T \rangle \, \Gamma \, \bar{\psi} (i\partial)^k \chi \,, \tag{1}$$

where k is an integer power, λ is a coupling constant, and ψ and χ are fermion fields. In this expression, the tensor vev $\langle T \rangle$ carries spacetime indices, which are suppressed in this notation. This vev is effectively a set of functions or constants that are fixed in a given observer frame. What this means is that interactions with these coefficients can have preferred directions in spacetime or velocity (boost) dependence. The vev coefficients therefore induce Lorentz violation.

Note that the higher-dimensional ($k > 0$) derivative couplings are expected to be balanced by additional inverse factors of a large mass scale, which is assumed to be the Planck mass m_P. In a more complete low-energy effective theory describing fermions ψ and χ there could also be other terms with additional couplings, including possible Yukawa couplings. A more general interaction term of the form in (1) at order k could then be written as

$$\mathcal{L} \sim t^{(k)} \, \Gamma \, \bar{\psi} (i\partial)^k \chi \,, \tag{2}$$

where the coefficient $t^{(k)}$, which carries spacetime indices, absorbs all of the couplings, inverse mass factors, and the vev. This effective coefficient acts essentially as a fixed background field that induces Lorentz violation. In addition to interactions with fermions, additional terms involving gauge-field couplings and gravitational interactions are possible as well. A generalization would be to include all possible contractions of known SM and gravitational fields with fixed background coefficients $t^{(k)}$.

This generalization to include all arbitrary-dimension interaction terms inducing Lorentz violation in effective field theory is the idea behind the SME [7–9]. Note as well that each term is assumed to be an observer scalar, with all spacetime indices contracted. The full SME is then defined as the effective field theory obtained when all such scalar terms are formed using SM and gravitational fields contracted with coefficients that induce Lorentz violation. The SME coefficients (the generalized $t^{(k)}$ factors) are assumed to be heavily suppressed, presumably by inverse powers of the Planck mass. The extent of the suppression increases with order k. Without a completely viable string field theory, it is not possible to assign definite numerical values to these coefficients, and clearly (as in the SM itself) there are hierarchy issues. However, since no Lorentz violation has been observed in nature, it must be that the SME coefficients are small. Alternatively, one can adopt a phenomenological approach and treat the coefficients as quantities to be bounded in experiments. Such bounds will also constitute a measure of the sensitivity to Lorentz violation attained in the experiment.

Interestingly, although the SME was originally motivated from ideas in string field theory, including the idea of spontaneous Lorentz symmetry breaking, its relevance and usefulness extend well beyond these ideas. In fact, there is nothing in the SME that requires that the Lorentz-violation coefficients emerge due to a process of spontaneous Lorentz violation. The SME coefficients can also be viewed as due to explicit Lorentz violation or as arising from some unknown mechanism. Indeed, once the philosophy of the SME is appreciated – that it is the most general observer-independent field theory incorporating Lorentz violation – then no matter what scalar Lagrangian is written down involving known low-energy fields, the result will be contained in the full SME.

An illustration of this comes from studying noncommutative field theory. These are theories that have noncommuting coordinates

$$[x^\mu, x^\nu] = i\theta^{\mu\nu} \,. \tag{3}$$

It has been shown that this type of geometry can occur naturally in string theory [15], and that it leads to Lorentz violation [16]. Here, however, the mechanism leading to Lorentz violation is in general different from that of spontaneous symmetry breaking. Nonetheless, the form of the effective interactions that arise are fully contained in the SME. The fixed parameters $\theta^{\mu\nu}$, which break the Lorentz symmetry, act effectively so as to produce SME coefficients. For example, the effective field theory involving a $U(1)$ gauge field in a noncommutative geometry includes lagrangian terms of the form

$$\mathcal{L} \sim \frac{1}{4}\, iq\, \theta^{\alpha\beta}\, F_{\alpha\beta}\, \bar{\psi}\, \gamma^\mu\, D_\mu\, \psi \,, \tag{4}$$

where $F_{\alpha\beta}$ is the field strength. Here, as in (1) the interaction takes the form of a scalar-valued product of known particle fields, derivative operators, and a set of fixed background functions.

There are a number of other examples of effective field theories with Lorentz violation that have been put forward in recent years, with a wide variety of

motivations or ideas for symmetry breaking. Nonetheless, as long as the resulting theories are described by scalar lagrangians, then they are compatible with the approach of the SME. For example, a model with spatial rotational invariance was used in [17] to study high-energy cosmic rays above the GZK cutoff. Another example with a higher-dimensional lagrangian giving rise to Lorentz-violating dispersion relations was considered in [18]. An example involving gravitational fields includes a parameterized set of kinetic terms for a vector field in a theory with spontaneous Lorentz breaking [19]. In all of these cases, the lagrangian terms can be found as a subset of the full SME.

Over the years, a number of phenomenological frameworks that involve specific types of Lorentz violation have been developed and used extensively by experimentalists. A sampling includes the $TH\epsilon\mu$ model [20], the Robertson-Mansouri-Sexl framework [21], the PPN formalism [22], as well as models based on kinematical breaking of Lorentz symmetry (see [12, 14] for reviews). In some cases, these theories describe parameterized equations of motion or dispersion relations and do not originate from a scalar lagrangian. However, to the extent that these models can be described by effective field theory defined by a scalar lagrangian, they are compatible with the SME and direct links between their parameterizations and the SME coefficients can be obtained.

It should be noted as well, that in addition to breaking Lorentz symmetry, the SME also leads to violation of the discrete symmetry CPT [4,5]. This symmetry is the product of charge conjugation (C), parity (P), and time reversal (T). According to the CPT theorem [23], a relativistic field theory describing point particles should exactly obey CPT symmetry. Conversely, a second theorem states that if CPT is violated in field theory, then Lorentz symmetry must also be broken [24]. It then follows that any observer-independent effective field theory describing CPT violation must also be contained within the SME. Since CPT can be tested to very high precision in experiments with matter and antimatter, this opens up a whole new avenue for exploring the phenomenology of Lorentz violation.

In summary, the full SME is defined as the most general observer-independent theory of Lorentz and CPT violation that contains the SM and gravity. It thus provides a unifying framework that can be used to investigate possible signals of Lorentz and CPT violation. Because it contains an infinity of terms, with an unlimited set of coefficients with spacetime indices, it is not possible to list all of them. However, the terms can be classified in a general way, and a uniform notation can be developed. It is also possible to develop subset theories of the full SME, which contain a finite number of terms. One subset in particular has been investigated extensively in recent experiments. It is the minimal SME, which is comprised of the gauge-invariant subset of terms in the full SME with dimension four or less.

Finally, one other remark about the SME coefficients is worth mentioning. It is often commented these coefficients, such as for example a nonzero vacuum vev of a tensor field generated from a process of spontaneous Lorentz violation, are reminiscent of the old pre-relativistic ether. However, the ether was thought to

be a medium (with a rest frame) for light, whereas an SME coefficient need not be thought of in this way. The SME coefficients act effectively as background vacuum fields. Their interactions typically select out a particular particle species. In fact, if that particle is not the photon, then the SME coefficient will have no direct influence on the speed of light. Moreover, the SME coefficients carry tensor indices and therefore have definite spacetime directions in any observer frame. In the end, while there are some similarities to the old ether, the physical effects of the SME coefficients are significantly different.

3 Constructing the SME

One of the defining features of the SME is that the theory is observer independent [8]. It is therefore important to make clear the distinction between what are called *observer* and *particle* Lorentz transformations. An observer Lorentz transformation is a change of observer frame. It can be viewed as a rotation or boost of the basis vectors in the local frame. The philosophy of the SME is that even with Lorentz violation, physics must remain observer independent. The results of an experiment should not depend on the chosen perspective of any observer. In contrast, a particle Lorentz transformation is a rotation or boost performed on an individual particle field while leaving the coordinate frame fixed. In this case, if there is Lorentz violation, the physics can change.

In terms of what this means for an experiment, the observer invariance of the SME says that the results of a measurement cannot depend on the choice of coordinate frame or observational perspective made by the experimenter. On the other hand, if Lorentz symmetry is broken, the results of the experiment can change if the apparatus itself is rotated or boosted in some direction, both of which are examples of particle Lorentz transformations.

Note that this feature of the SME breaks the relativity principle, which is a central assumption of (unbroken) relativity theory. This principle is often stated as the equivalence of passive and active Lorentz transformations when one is performed as the inverse of the other. In the formulation of the SME, however, the terms passive and active are deliberately avoided since for one thing their usage is sometimes confused in the literature. More importantly, though, it is observer independence that is the physically defining feature of the theory, and the terminology should reflect this. In addition, observers need not be inactive. The idea in the SME is that even if an observer actively changes its perspective or relative motion with respect to the apparatus in an experiment, the results of measurements should remain unchanged.

A similar distinction between observer and particle transformations can be made for general coordinate transformations performed in the spacetime manifold of a Riemann or Riemann-Cartan geometry [9, 10]. An observer transformation is simply a change of spacetime coordinates, which leaves the physics unchanged. On the other hand, a particle transformation is essentially a diffeomorphism, which maps one point on the spacetime to another. The change in a tensor under pullback to the original point is given by the Lie derivative.

The full SME is defined using a vierbein formalism. This permits a natural distinction between the spacetime manifold and local Lorentz frames. The vierbein ${e_\mu}^a$ provides a link between the components of a tensor field $T_{\lambda\mu\nu\ldots}$ on the spacetime manifold (denoted using Greek indices) and the corresponding components $T_{abc\ldots}$ in a local Lorentz frame (denoted using latin indices). The correspondence is given by

$$T_{\lambda\mu\nu\ldots} = {e_\lambda}^a {e_\mu}^b {e_\nu}^c \cdots T_{abc\ldots} \,. \tag{5}$$

In this notation, the components of the spacetime metric are $g_{\mu\nu}$, while in a local Lorentz frame, the metric takes the Minkowski form η_{ab}. A necessary condition for the vierbein is therefore that $g_{\mu\nu} = {e_\mu}^a {e_\nu}^b \eta_{ab}$. Covariant derivatives acting on tensor fields with local indices introduce the spin connection ${\omega_\mu}^{ab}$. For example,

$$D_\mu {e_\nu}^a = \partial_\mu {e_\nu}^a - {\Gamma^\alpha}_{\mu\nu} {e_\alpha}^a + {\omega_\mu}^a{}_b {e_\nu}^b \,. \tag{6}$$

In a Riemann spacetime where $D_\lambda g_{\mu\nu} = 0$, the spin connection is not an independent field, but rather is a prescribed function of the vierbein and its derivatives. However, in a Riemann-Cartan spacetime the spin connection represents independent degrees of freedom associated with there being nonzero torsion.

The observer independence of the SME requires that all of the terms in the lagrangian be observer scalars under both general coordinate transformations and local Lorentz transformations. This means that every spacetime index and every local Lorentz index must be fully contracted in the lagrangian.

However, the SME is not invariant under particle diffeomorphisms and local Lorentz transformations. Explicitly, a diffeomorphism maps one point on the spacetime to another. It can be characterized infinitesimally in a coordinate basis by the transformation

$$x^\mu \to x^\mu + \xi^\mu \,. \tag{7}$$

The four infinitesimal parameters ξ^μ comprise the diffeomorphism degrees of freedom. On the other hand, under an infinitesimal particle Lorentz transformation the field components transform through contraction with a matrix of the form

$${\Lambda^a}_b \approx {\delta^a}_b + {\epsilon^a}_b \,, \tag{8}$$

where $\epsilon_{ab} = -\epsilon_{ba}$ are the infinitesimal parameters carrying the six Lorentz degrees of freedom and generating the local Lorentz group. Evidently, there are a total of ten relevant spacetime symmetries.

Violation of these symmetries occurs when an interaction term contains coefficients that remain fixed under a particle transformation, such as when as a particle rotation or boost is performed in a background with a fixed vev.

3.1 Minimal SME

The full SME consists of an unlimited number of observer scalar terms consisting of contractions of SM fields, gravitational fields, and SME coefficients. To begin

to explore phenomenology, it makes sense to advance incrementally. Since gauge symmetry and renormalizability are foundations of our current understanding in particle physics, a first increment would be to construct a subset theory that maintains these features. It is referred to as the minimal SME. It will first be defined in Minkowski spacetime and then generalized to include gravitational fields in a Riemann-Cartan geometry.

The minimal SME, constructed from dimension four or fewer operators, describes the leading-order effects of Lorentz violation. This is because the higher-dimensional terms are expected to be suppressed by additional inverse powers of the Planck mass compared to those in the minimal SME. Effects involving couplings to gravitational fields are also expected to be smaller than those involving interactions in the SM, particularly electrodynamic interactions. For this reason, the Lorentz tests described later on are investigated using primarily a QED subset of the minimal SME in flat Minkowski spacetime. Nonetheless, it should be kept in mind that a particular type of Lorentz violation might only occur at subleading order. For this reason, it is important ultimately to investigate more general tests in the context of the full SME, including gravitational effects as well as interactions involving higher-dimensional terms. However, that goes beyond the scope of this overview.

To construct the minimal SME in flat Minkowski spacetime [8], the first ingredient that must be put in is the minimal SM itself. This consists of quark and lepton sectors, gauge fields, and a Higgs sector. Denote the left- and right-handed lepton and quark multiplets by

$$L_A = \begin{pmatrix} \nu_A \\ l_A \end{pmatrix}_L , \qquad R_A = (l_A)_R , \tag{9}$$

$$Q_A = \begin{pmatrix} u_A \\ d_A \end{pmatrix}_L , \quad U_A = (u_A)_R , \quad D_A = (d_A)_R , \tag{10}$$

where $A = 1, 2, 3$ labels the flavor, with $l_A = (e, \mu, \tau)$, $\nu_A = (\nu_e, \nu_\mu, \nu_\tau)$, $u_A = (u, c, t)$, and $d_A = (d, s, b)$. The Higgs doublet is denoted by ϕ. The SU(3), SU(2), and U(1) gauge fields are G_μ, W_μ, and B_μ, respectively, with corresponding field strengths: $G_{\mu\nu}$, $W_{\mu\nu}$, and $B_{\mu\nu}$. The gauge couplings are g_3, g, and g', while q denotes the electric charge. The Yukawa couplings are G_L, G_U, G_D.

The relevant sectors in the SM Lagrangian are:

$$\mathcal{L}_{\text{lepton}} = \frac{1}{2} i \bar{L}_A \gamma^\mu \overleftrightarrow{D}_\mu L_A + \frac{1}{2} i \bar{R}_A \gamma^\mu \overleftrightarrow{D}_\mu R_A \tag{11}$$

$$\mathcal{L}_{\text{quark}} = \frac{1}{2} i \bar{Q}_A \gamma^\mu \overleftrightarrow{D}_\mu Q_A + \frac{1}{2} i \bar{U}_A \gamma^\mu \overleftrightarrow{D}_\mu U_A + \frac{1}{2} i \bar{D}_A \gamma^\mu \overleftrightarrow{D}_\mu D_A \tag{12}$$

$$\mathcal{L}_{\text{Yukawa}} = -\left[(G_L)_{AB} \bar{L}_A \phi R_B + (G_U)_{AB} \bar{Q}_A \phi^c U_B + (G_D)_{AB} \bar{Q}_A \phi D_B \right] , \tag{13}$$

$$\mathcal{L}_{\text{Higgs}} = (D_\mu \phi)^\dagger D^\mu \phi + \mu^2 \phi^\dagger \phi - \frac{\lambda}{3!} (\phi^\dagger \phi)^2 \tag{14}$$

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{2} \text{Tr}(G_{\mu\nu} G^{\mu\nu}) - \frac{1}{2} \text{Tr}(W_{\mu\nu} W^{\mu\nu}) - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} , \tag{15}$$

where D_μ are gauge-covariant derivatives.

The SME introduces additional Lagrangian terms that are contractions of these SM fields with the SME coefficients. The SME coefficients are constrained by the requirement that the Lagrangian be hermitian. For an SME coefficient with an even number of spacetime indices, the pure trace component is irrelevant because it maintains Lorentz invariance. Such coefficients may therefore be taken as traceless.

In the fermion sector of the minimal SME, four sets of terms can be classified according to whether they involve leptons or quarks and whether CPT is even or odd. They are

$$\mathcal{L}^{\text{CPT}-\text{even}}_{\text{lepton}} = \frac{1}{2} i (c_L)_{\mu\nu AB} \bar{L}_A \gamma^\mu \overleftrightarrow{D^\nu} L_B + \frac{1}{2} i (c_R)_{\mu\nu AB} \bar{R}_A \gamma^\mu \overleftrightarrow{D^\nu} R_B \tag{16}$$

$$\mathcal{L}^{\text{CPT}-\text{odd}}_{\text{lepton}} = -(a_L)_{\mu AB} \bar{L}_A \gamma^\mu L_B - (a_R)_{\mu AB} \bar{R}_A \gamma^\mu R_B \tag{17}$$

$$\begin{aligned} \mathcal{L}^{\text{CPT}-\text{even}}_{\text{quark}} = {} & \frac{1}{2} i (c_Q)_{\mu\nu AB} \bar{Q}_A \gamma^\mu \overleftrightarrow{D^\nu} Q_B + \frac{1}{2} i (c_U)_{\mu\nu AB} \bar{U}_A \gamma^\mu \overleftrightarrow{D^\nu} U_B \\ & + \frac{1}{2} i (c_D)_{\mu\nu AB} \bar{D}_A \gamma^\mu \overleftrightarrow{D^\nu} D_B \end{aligned} \tag{18}$$

$$\begin{aligned} \mathcal{L}^{\text{CPT}-\text{odd}}_{\text{quark}} = {} & -(a_Q)_{\mu AB} \bar{Q}_A \gamma^\mu Q_B - (a_U)_{\mu AB} \bar{U}_A \gamma^\mu U_B \\ & - (a_D)_{\mu AB} \bar{D}_A \gamma^\mu D_B \,. \end{aligned} \tag{19}$$

In these expressions, the coefficients a_μ have dimensions of mass, while $c_{\mu\nu}$ are dimensionless and traceless.

The couplings between fermions and the Higgs field are all CPT even and are

$$\begin{aligned} \mathcal{L}^{\text{CPT}-\text{even}}_{\text{Yukawa}} = -\frac{1}{2} \big[& (H_L)_{\mu\nu AB} \bar{L}_A \phi \sigma^{\mu\nu} R_B + (H_U)_{\mu\nu AB} \bar{Q}_A \phi^c \sigma^{\mu\nu} U_B \\ & + (H_D)_{\mu\nu AB} \bar{Q}_A \phi \sigma^{\mu\nu} D_B \big] \,, \end{aligned} \tag{20}$$

where the SME coefficients $H_{\mu\nu}$ are dimensionless and antisymmetric.

The Higgs sector itself can be CPT even or odd. The terms are

$$\begin{aligned} \mathcal{L}^{\text{CPT}-\text{even}}_{\text{Higgs}} = {} & \frac{1}{2} (k_{\phi\phi})^{\mu\nu} (D_\mu \phi)^\dagger D_\nu \phi - \frac{1}{2} (k_{\phi B})^{\mu\nu} \phi^\dagger \phi B_{\mu\nu} \\ & - \frac{1}{2} (k_{\phi W})^{\mu\nu} \phi^\dagger W_{\mu\nu} \phi \,, \end{aligned} \tag{21}$$

$$\mathcal{L}^{\text{CPT}-\text{odd}}_{\text{Higgs}} = i (k_\phi)^\mu \phi^\dagger D_\mu \phi \,. \tag{22}$$

The dimensionless coefficient $k_{\phi\phi}$ can have symmetric real and antisymmetric imaginary parts. The other coefficients have dimensions of mass.

The gauge sector consists of

$$\mathcal{L}_{\text{gauge}}^{\text{CPT-even}} = -\frac{1}{2}(k_G)_{\kappa\lambda\mu\nu}\text{Tr}(G^{\kappa\lambda}G^{\mu\nu}) - \frac{1}{2}(k_W)_{\kappa\lambda\mu\nu}\text{Tr}(W^{\kappa\lambda}W^{\mu\nu}) - \frac{1}{4}(k_B)_{\kappa\lambda\mu\nu}B^{\kappa\lambda}B^{\mu\nu}\,, \tag{23}$$

$$\begin{aligned}\mathcal{L}_{\text{gauge}}^{\text{CPT-odd}} &= (k_3)_\kappa\epsilon^{\kappa\lambda\mu\nu}\text{Tr}(G_\lambda G_{\mu\nu} + \frac{2}{3}ig_3G_\lambda G_\mu G_\nu) \\ &\quad +(k_2)_\kappa\epsilon^{\kappa\lambda\mu\nu}\text{Tr}(W_\lambda W_{\mu\nu} + \frac{2}{3}igW_\lambda W_\mu W_\nu) \\ &\quad +(k_1)_\kappa\epsilon^{\kappa\lambda\mu\nu}B_\lambda B_{\mu\nu} + (k_0)_\kappa B^\kappa\,.\end{aligned} \tag{24}$$

The coefficients $k_{G,W,B}$ are dimensionless, have the symmetries of the Riemann tensor, and have a vanishing double trace. The coefficients $k_{1,2,3}$ are real and have dimensions of mass, while k_0 is also real and has dimensions of mass cubed. Note that if any of these CPT-odd terms appear in the theory, they would generate instabilities associated with negative contributions to the energy. For this reason, the coefficients $k_{0,1,2,3}$ are assumed to vanish. Interestingly, it appears that no radiative corrections in the SME appear to generate nonzero values for these coefficients, at least to one loop.

It is also important to realize that some of the SME terms can be eliminated by field redefinitions [8,25,26]. For example, some of the terms involving the coefficients $a_{L,R,Q,U,D}$ can be eliminated by position-dependent field-phase redefinitions. Another example is that certain terms involving the coefficients $c_{L,R,Q,U,D}$ can be absorbed by the terms involving the coefficients $H_{L,U,D}$ through field-normalization redefinitions. In particular, what this means is that while a field theory can be written down that ostensibly has explicit Lorentz violation, it is sometimes the case that there are no physical effects because the theory is equivalent through field redefinitions to a Lorentz-invariant theory.

Clearly, there are a number of additional theoretical issues that are relevant to the construction of the SME as a consistent low-energy field theory incorporating Lorentz violation. These include a more in-depth discussion of the nature of field theory with Lorentz violation (including quantization of the theory) [8], issues related to causality [25], the possibility of additional extensions including for example supersymmetry [27], renormalization [28], electroweak symmetry breaking [8], radiative corrections [29], spacetime variations of couplings [30], etc. It is not possible to describe all of these issues here. The interested reader is referred to the original papers.

3.2 Gravity Sector

The gravity sector of the SME has been discussed in [9], and the minimal theory (dimension four or fewer terms) has been explicitly constructed. A vierbein formalism is used, which gives the theory a close parallel to gauge theory. Lorentz breaking occurs due to the presence of SME coefficients, which remain fixed under particle Lorentz transformations in a local frame. In this case, the SME coefficients carry Latin indices, e.g., b_a for a vector, with respect to the local basis set. The conversion to spacetime coordinates is implemented by the vierbein,

giving, e.g., $b_\mu = e_\mu{}^a b_a$. The lagrangian can then be written in terms of fields and SME coefficients defined on the spacetime manifold. A natural (though not required) assumption is that the SME coefficients are smooth functions over the manifold. It is not necessary to require that they be covariantly constant. In fact, defining covariantly constant tensors over a manifold places stringent topological constraints on the geometry. One simplifying assumption, which could occur naturally in the context of spontaneous Lorentz breaking, is to assume that the SME coefficients are constants in the local frame. However, again, this is not a requirement in the formulation of the SME theory.

To construct the minimal SME including gravity, the first step is to incorporate gravitational fields into the usual SM. This is done by rewriting all of the terms in (11) through (15) with fields and gamma matrices defined with respect to the local frame (using Latin indices). The vierbein is then used to convert these terms over to the spacetime manifold. Factors of the determinant of the vierbein e are included as well so that integration of the lagrangian density (giving the action) is covariant. Derivatives are understood as well to be both spacetime and gauge covariant. With these changes, (11), for example, becomes

$$\mathcal{L}_{\text{lepton}} = \frac{1}{2} i e e^\mu{}_a \bar{L}_A \gamma^a \overleftrightarrow{D}_\mu L_A + \frac{1}{2} i e e^\mu{}_a \bar{R}_A \gamma^a \overleftrightarrow{D}_\mu R_A \,. \tag{25}$$

The other terms for the quark, Yukawa, Higgs, and gauge sectors follow a similar pattern.

The Lorentz-violating SME terms constructed from SM fields are obtained in a similar way. The various particle sectors can again be divided between CPT odd and even contributions. Each of the terms in (16) to (24) is then written using local indices and vierbeins, which convert the equations over to the spacetime manifold. As an example, (16) becomes

$$\mathcal{L}_{\text{lepton}}^{\text{CPT}-\text{even}} = -\frac{1}{2} i (c_L)_{\mu\nu AB} e e^\mu{}_a \bar{L}_A \gamma^a \overleftrightarrow{D}^\nu L_B - \frac{1}{2} i (c_R)_{\mu\nu AB} e e^\mu{}_a \bar{R}_A \gamma^a \overleftrightarrow{D}^\nu R_B \,. \tag{26}$$

The remaining equations follow the same pattern.

The pure-gravity sector of the minimal SME consists of a Lorentz-invariant gravity sector and a Lorentz-violating sector. The Lorentz-invariant lagrangian consists of terms that are products of the gravitational fields. In the general case, this includes terms constructed from curvature, torsion, and covariant derivatives. Einstein's gravity (with or without a cosmological term) would be a special case in this sector.

The Lorentz-violating Lagrangian terms in the gravity sector of the minimal SME are constructed by combining the SME coefficients with gravitational field operators to produce an observer scalar under local Lorentz transformations and general coordinate transformations. These consist of products of the vierbein, the spin connection, and their derivatives, but for simplicity they can be written in terms of the curvature, torsion, and covariant derivatives. The minimal case (up to dimension four) has the form:

$$\mathcal{L}_{e,\omega}^{\mathrm{LV}} = e(k_T)^{\lambda\mu\nu} T_{\lambda\mu\nu} + e(k_R)^{\kappa\lambda\mu\nu} R_{\kappa\lambda\mu\nu} + e(k_{TT})^{\alpha\beta\gamma\lambda\mu\nu} T_{\alpha\beta\gamma} T_{\lambda\mu\nu} + e(k_{DT})^{\kappa\lambda\mu\nu} D_\kappa T_{\lambda\mu\nu} \,. \tag{27}$$

The SME coefficients in this expression have the symmetries of the associated Lorentz-violating operators. All except $(k_T)^{\lambda\mu\nu}$, which has dimensions of mass, are dimensionless.

The Lorentz-violating sector introduces additional gravitational couplings that can have phenomenological consequences, including effects on cosmology, black holes, gravitational radiation, and post-Newtonian physics. As a starting point for a phenomenological investigation of the gravitational consequences of Lorentz violation, it is useful to write down the Riemannian limit of the minimal SME gravity sector. It is given as [9]

$$S_{e,\omega,\Lambda} = \frac{1}{2\kappa} \int d^4x \left[e(1-u)R - 2e\Lambda + e s^{\mu\nu} R_{\mu\nu} + e t^{\kappa\lambda\mu\nu} R_{\kappa\lambda\mu\nu} \right] . \tag{28}$$

The SME coefficient $(k_R)^{\kappa\lambda\mu\nu}$ has been expanded into coefficients $s^{\mu\nu}$, $t^{\kappa\lambda\mu\nu}$, and u that distinguish the effects involving the Riemann, Ricci, and scalar curvatures. The coefficients ß have the symmetries of the Ricci tensor, while $t^{\kappa\lambda\mu\nu}$ has those of the Riemann tensor. Taking tracelessness conditions into account, there are 19 independent components.

Another useful limit is the QED subset of the SME. This extension in Minkowski space has been used extensively to investigate high-precision experimental tests of Lorentz symmetry in atomic and particle systems. Generalizing to include gravity involves introducing additional vierbein-fermion couplings as well as a pure-gravity sector. These additional terms can then be investigated for potential signals of Lorentz violation due to gravitational effects in high-precision experiments.

A full treatment of the gravity sector of the SME should include looking at the energy momentum tensor, Einstein's equations, and consistency relations between these stemming from, for example, the Bianchi identities. These types of issues are described in depth in [9]. Interestingly, a difference between theories with explicit versus spontaneous breaking of Lorentz symmetry is found. In a generic Riemann-Cartan theory with explicit breaking of Lorentz symmetry, the Bianchi identities are not consistent with the the covariant conservation laws and equations of motion. On the other hand, if Lorentz symmetry is spontaneously broken, the problem is evaded.

4 Spontaneous Lorentz Violation

One of the original motivations for developing the SME was that mechanisms in string theory suggest that local Lorentz symmetry might be spontaneously broken [4]. While the full SME describes any observer-independent Lorentz violation at the level of effective field theory, one important special case is when Lorentz symmetry is spontaneously broken. This provides an elegant mechanism

in which the symmetry holds dynamically, but is broken (or hidden) by the solutions of the theory. The lagrangian and equations of motion still respect the symmetry, however; the vacuum values of the fields do not. Tensor-valued fields acquire nonzero vevs which have definite spacetime directions, thereby breaking the symmetry under boosts and rotations.

There are certain theoretical issues that arise when the Lorentz violation is due to spontaneous symmetry breaking. This section examines some of these issues, in particular, what the fate is of the Nambu-Goldstone modes when Lorentz symmetry is spontaneously broken.

In gauge theory, it is well known that when a continuous global symmetry is spontaneously broken, massless Nambu-Goldstone (NG) modes appear [31]. If instead the broken symmetry is local, then a Higgs mechanism can occur in which the gauge bosons become massive [32]. The question naturally arises as to what the fate of the NG modes is when Lorentz symmetry is spontaneously broken and whether a Higgs mechanism can occur for the case of local Lorentz symmetry (as in a theory with gravity).

This question has recently been addressed in detail in [10]. A generic analysis of theories with spontaneous Lorentz breaking was carried out in Riemann-Cartan spacetime and in the limiting cases of Riemann and Minkowski spacetime. A number of general features were found.

First, a connection between spontaneous breaking of local Lorentz symmetry and diffeomorphisms was found to hold. This occurs because when the vierbein takes a vacuum value, which for simplicity we can take as its value in a Minkowski background, $e_\mu{}^a = \delta^a_\mu$, then if a local tensor acquires a fixed vev, e.g., b_a for the case of a vector, which breaks local Lorentz symmetry, then the associated spacetime vector b_μ as given by contraction with the vierbein also acquires a fixed vev. The spacetime vev b_μ breaks diffeomorphisms. The converse is also true. If a nonscalar tensor vev on the spacetime manifold breaks diffeomorphisms, then the associated local tensor will have a vev that breaks local Lorentz symmetry. In the case of a scalar, the derivatives of the field will have vevs that break local Lorentz symmetry.

Next, the question of how many NG modes there are and where they reside was examined. Since there are six Lorentz symmetries and four diffeomorphisms, which can all be broken when a tensor with a sufficient number of indices acquires a fixed vev, this means that in general up to ten NG modes can appear. A general argument shows that these ten modes can all be absorbed as additional degrees of freedom in the vierbein. A simple counting argument supports this as well. The vierbein has 16 components. With Lorentz symmetry, six of these modes can be gauged away. They are usually chosen as the antisymmetric components. Similarly, diffeomorphisms can be used to remove four additional degrees of freedom. This leaves six vierbein modes in the general case. Einstein's theory has four of these modes as auxiliary, resulting in only two massless modes for the graviton. However, in a more general gravitational theory, there can be up to six propagating modes, which in a vierbein formalism are the six vierbein degrees of freedom. If Lorentz symmetry and diffeomorphisms are broken, then the ability

to gauge away some of the vierbein degrees of freedom is lost. In particular, since up to ten symmetries can be broken, up to ten additional modes (the NG modes) can appear in the vierbein.

The number of NG modes is also affected by the nature of the vev and by the fact that the symmetry is a spacetime symmetry. For example, in the case of a vector vev, which breaks three Lorentz symmetries and one diffeomorphism, it might be expected that there would be three massless NG Lorentz modes and one massless NG diffeomorphism mode. However, in the case where the vector vev is a constant, the diffeomorphism mode is found to be an auxiliary mode. It is also found that there are only two propagating massless Lorentz modes. The third Lorentz mode is found to be auxiliary as well. In this case, since the NG modes carry vector indices, it makes sense that a massless vector would only have two propagating modes. This clearly provides an example where the usual counting of NG modes (one massless mode per broken generator) does not hold for the case of a broken spacetime symmetry [33].

It was also found that the fate of the NG modes depends on the geometry. In Riemann or Minkowski spacetime, where the torsion is zero, the NG modes appear as additional massless or auxiliary modes in the vierbein. However, in Riemann-Cartan spacetime, which has nonzero torsion and where the spin connection has degrees of freedom that are independent from the vierbein, the possibility of a Higgs mechanism occurs. This is because a mass term for the spin connection can form when local Lorentz symmetry is spontaneously broken. If the theory permits massless propagating modes for the spin connection, then these modes can acquire a mass. In principle, the mechanism is straightforward. However, finding a ghost-free unbroken model with a propagating spin connection that is compatible with the mass term is challenging.

A specific vector model with spontaneous Lorentz breaking, called a bumblebee model, has been used to illustrate the behavior of the NG modes. For simplicity, this overview will concentrate entirely on this example for the case of a constant vev. All of the general features described above will be applicable.

Bumblebee models in a gravitational theory were first looked at by Kostelecky and Samuel as a simple model for investigating the consequences of spontaneous Lorentz violation [11]. Their properties have been studied in a variety of contexts [34]. Much of the attention has focused on models with a timelike vev. It has been suggested that if a NG diffeomorphism mode propagates in this case, then it would have an unusual dispersion relation [35].

One especially noteworthy feature of the bumblebee model occurs in Minkowski and Riemann spacetime. It is found (in the linearized theory) that the massless NG Lorentz modes behave essentially as the photon in an axial gauge [10]. Connections between Lorentz breaking and gauge fixing have been noted previously, leading to the suggestion that the photon is comprised of NG modes due to spontaneous Lorentz breaking [36,37]. However, the approach of the bumblebee model is different. It is not a $U(1)$ gauge theory, since it contains a potential V that is not $U(1)$ invariant. The Lorentz breaking is therefore not a $U(1)$ gauge fixing choice. Nonetheless, the NG modes appear to behave at lowest order as photons

in an axial gauge. Moreover, there are additional tell-tale signs of Lorentz breaking [10]. These include additional SME couplings in Riemann and Minkowski spacetime as well as anomalous gravitational couplings in the case of a Riemann geometry. This offers the possibility of letting experiments determine whether massless photons are the result of unbroken gauge symmetry or whether they might be due to spontaneously broken Lorentz symmetry.

4.1 Bumblebee Models

The definition of a bumblebee model is that it is a vector theory in which the vector field B^μ acquires a nonzero vev, which spontaneously breaks Lorentz symmetry. The lagrangian consists of a kinetic term for B^μ and a potential V that induces spontaneous Lorentz breaking. The potential is not $U(1)$ gauge invariant. Typically, the potential imposes a vev $b_a \neq 0$ for the vector in a local frame. The vierbein relates this back to the spacetime vector as $B_\mu = e_\mu{}^a b_a$. For simplicity, we assume a perturbative solution about a Minkowski background. This permits us to drop the distinction between latin and Greek indices and to write

$$e_{\mu\nu} = \eta_{\mu\nu} + \left(\frac{1}{2}h_{\mu\nu} + \chi_{\mu\nu}\right) , \tag{29}$$

where the ten symmetric excitations $h_{\mu\nu} = h_{\nu\mu}$ are associated with the metric $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$, while the six antisymmetric components $\chi_{\mu\nu} = -\chi_{\nu\mu}$ are the local Lorentz degrees of freedom. In this background, the vacuum solution takes the form

$$\langle B^\mu \rangle = b^\mu , \qquad \langle e_{\mu\nu} \rangle = \eta_{\mu\nu} . \tag{30}$$

There are a number of choices for the kinetic and potential terms. Vector-current interactions and additional vector-curvature couplings that are forbidden in $U(1)$ gauge theory can be included as well [4, 9].

Here, as an illustrative example, we examine the model given by the Lagrangian

$$\begin{aligned} \mathcal{L}_B = \frac{1}{2\kappa}(eR + \xi e B^\mu B^\nu R_{\mu\nu}) - \frac{1}{4}eB_{\mu\nu}B^{\mu\nu} \\ -e\lambda(B_\mu B^\mu \pm b^2) - eB_\mu J^\mu , \end{aligned} \tag{31}$$

where $\kappa = 8\pi G$ and ξ is a coupling coefficient between the vector field and the curvature. The kinetic terms in this example are analogous to those in Einstein-Maxwell theory. However, in the general case in a Riemann-Cartan spacetime, the torsion contributes to these terms and the field strength is defined by

$$B_{\mu\nu} = D_\mu B_\nu - D_\nu B_\mu , \tag{32}$$

where D_μ are covariant derivatives. The potential term is

$$V(B_\mu B^\mu \pm b^2) = \lambda(B_\mu B^\mu \pm b^2) , \tag{33}$$

where λ is a Lagrange-multiplier field. It imposes the constraint that the vector field has a vev b^a obeying $b_a b^a = \mp b^2$ (with the sign corresponding to whether the vector is timelike or spacelike). The vector field can then be written in terms of the vierbein and can be expanded perturbatively to give

$$B^\mu = e^\mu{}_a b^a \approx b^\mu + \left(-\frac{1}{2}h^{\mu\nu} + \chi^{\mu\nu}\right) b_\nu \,. \tag{34}$$

The vierbein degrees of freedom include the NG modes.

This model can be studied in a linearized approximation. The symmetric and antisymmetric components of the vierbein transform as

$$\begin{aligned} h_{\mu\nu} &\to h_{\mu\nu}, \\ \chi_{\mu\nu} &\to \chi_{\mu\nu} - \epsilon_{\mu\nu} \,, \end{aligned} \tag{35}$$

under infinitesimal Lorentz transformations, while under infinitesimal diffeomorphisms

$$\begin{aligned} h_{\mu\nu} &\to h_{\mu\nu} - \partial_\mu \xi_\nu - \partial_\nu \xi_\mu, \\ \chi_{\mu\nu} &\to \chi_{\mu\nu} - \frac{1}{2}(\partial_\mu \xi_\nu - \partial_\nu \xi_\mu) \,. \end{aligned} \tag{36}$$

In these expressions, quantities of order (ϵh), $(\epsilon\chi)$, (ξh), $(\xi\chi)$, etc. are assumed small and hence negligible in the linearized treatment.

The NG modes can be found as the virtual fluctuations about the vacuum solution. These can be written as

$$\delta B^\mu = B^\mu - b^\mu \approx \left(-\frac{1}{2}h^{\mu\nu} + \chi^{\mu\nu}\right) b_\nu \,. \tag{37}$$

It is useful to introduce projections on the transverse and longitudinal components of δB^μ along b^μ. Assuming $b^2 \neq 0$, these are given by

$$(P_\parallel)^\mu{}_\nu = \frac{b^\mu b_\nu}{b^\sigma b_\sigma} \,, \quad (P_\perp)^\mu{}_\nu = \delta^\mu{}_\nu - (P_\parallel)^\mu{}_\nu \,. \tag{38}$$

Defining the projected fluctuations as

$$\mathcal{E}^\mu = (P_\perp)^\mu{}_\nu \delta B^\nu \,, \qquad \rho^\mu = (P_\parallel)^\mu{}_\nu \delta B^\nu \approx b^\mu \rho \,, \tag{39}$$

where

$$\rho = -\frac{b^\mu h_{\mu\nu} b^\nu}{2 b^\sigma b_\sigma} \,. \tag{40}$$

lets us write the field B^μ as

$$B^\mu \approx (1+\rho) b^\mu + \mathcal{E}^\mu \,. \tag{41}$$

In terms of these projections, the NG Lorentz and diffeomorphism modes can be identified. Under a virtual local particle Lorentz transformation only

components $\mathcal{E}^\mu$ obeying $b_\mu \mathcal{E}^\mu = 0$ are excited. These are the NG Lorentz modes, which evidently obey a condition similar to an axial-gauge condition in $U(1)$ gauge theory. If instead a virtual infinitesimal diffeomorphism is performed, only the longitudinal component ρ is excited. It can therefore be identified as the NG diffeomorphism mode. Note that a metric fluctuation about the vacuum solution,

$$\eta_{\mu\nu} \to g_{\mu\nu} \approx \eta_{\mu\nu} - \partial_\mu \xi_\nu - \partial_\nu \xi_\mu \,, \tag{42}$$

is generated by the diffeomorphism as well.

The dynamics of the NG modes depend on the background geometry. Three cases corresponding to Minkowski, Riemann, and Riemann-Cartan spacetime are examined in the following sections.

4.2 Minkowski Spacetime

In Minkowski spacetime, the curvature and torsion equal zero, and the metric can be written as

$$g_{\mu\nu} = \eta_{\mu\nu} \,. \tag{43}$$

The bumblebee Lagrangian in (31) reduces to

$$\mathcal{L}_B = -\frac{1}{4} B_{\mu\nu} B^{\mu\nu} - \lambda (B_\mu B^\mu \pm b^2) - B_\mu J^\mu \,. \tag{44}$$

In this case, it is found that the diffeomorpism mode ρ cancels in $B_{\mu\nu}$. It is therefore an auxiliary mode and does not propagate. The Lorentz modes are contained in the projection $\mathcal{E}_\mu$. Renaming this as $\mathcal{E}_\mu \equiv A_\mu$ and calling the field strength $F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu$ lets us rewrite the Lagrangian as

$$\mathcal{L}_B \to \mathcal{L}_{\rm NG} \approx -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - A_\mu J^\mu - b_\mu J^\mu + b^\mu \partial_\nu \Xi_\mu J^\nu \,, \tag{45}$$

where Ξ_μ is the longitudinal diffeomorphism mode ξ_μ promoted to an NG field. It is defined by $\rho = \partial_\mu \Xi^\mu$. Note that varying with respect to this auxiliary mode yields the current-conservation law, $\partial_\mu J^\mu = 0$.

The Lagrangian $\mathcal{L}_{\rm NG}$ is the effective quadratic lagrangian that governs the propagation of the NG modes in Minkowski space. The field A^μ has three degrees of freedom and automatically obeys an axial-gauge condition $b_\mu A^\mu = 0$. It contains the three Lorentz NG modes. Depending on the vev b_μ, the special cases of temporal gauge ($A^0 = 0$) and pure axial gauge ($A^3 = 0$) are possible.

It can be seen that in Minkowski spacetime the NG modes resemble those of a massless photon in $U(1)$ gauge theory in an axial gauge. Unlike the gauge theory case, however, where the masslessness of the photon is due to unbroken gauge symmetry, in this case the masslessness of the photon is a consequence of spontaneously broken Lorentz symmetry. An important question is whether this interpretation of the photon has experimentally verifiable consequences. Clearly, there is one additional interaction that does not hold for the usual photon in gauge theory. This is the Lorentz-violating term $b_\mu J^\mu$, where J^μ is the charge

current. This term can be identified with the SME term with coefficient a_μ^e that occurs in the QED limit of the SME [8]. This type of SME coefficient if it is constant is known to be unobservable in experiments restricted to the electron sector [8,9,25]. However, it can generate signals in the quark and neutrino sectors. Thus, in experiments with multiple particle sectors, the idea that the photon results from spontaneous Lorentz violation can potentially be tested in Minkowski space.

4.3 Riemann Spacetime

In Riemann geometry in a vierbein formalism, the spin connection $\omega_\mu{}^{ab}$ appears in covariant derivatives. However, the metric requirement,

$$D_\lambda e_\mu{}^a = 0\,, \tag{46}$$

and the fact that the torsion vanishes permits the spin connection to be completely determined in terms of the vierbein as

$$\begin{aligned}\omega_\mu{}^{ab} = {}& \frac{1}{2}e^{\nu a}(\partial_\mu e_\nu{}^b - \partial_\nu e_\mu{}^b) - \frac{1}{2}e^{\nu b}(\partial_\mu e_\nu{}^a - \partial_\nu e_\mu{}^a) \\ & -\frac{1}{2}e^{\alpha a}e^{\beta b}e_\mu{}^c(\partial_\alpha e_{\beta c} - \partial_\beta e_{\alpha c})\,.\end{aligned} \tag{47}$$

The spin connection has no independent degrees of freedom in Riemann spacetime, and the NG modes are still contained in the vierbein. In this case (with gravity), up to six of the 16 components of the vierbein can represent dynamical degrees of freedom associated with the gravitational fields.

We again consider the bumblebee lagrangian and vacuum as given in (31) and (30), respectively. The projector-operator decomposition of B^μ reveals that there are four potential NG modes contained in $\mathcal{E}^\mu$ and ρ, and the axial-gauge condition $b_\mu \mathcal{E}^\mu = 0$ still holds in Riemann spacetime. The field strength $B_{\mu\nu}$ can be rewritten as

$$B_{\mu\nu} = (\partial_\mu e_\nu{}^a - \partial_\nu e_\mu{}^a)b_a\,, \tag{48}$$

which suggests that the propagation of the vierbein is modified by the bumblebee kinetic term.

The effective lagrangian for the NG modes can be found by expanding the bumblebee lagrangian to quadratic order, keeping couplings to matter currents and curvature. The result in terms of the decomposed fields is

$$\begin{aligned}\mathcal{L}_{\rm NG} \approx {}& \frac{1}{2\kappa}\Big[eR + \xi e b^\mu b^\nu R_{\mu\nu} + \xi e A^\mu A^\nu R_{\mu\nu} \\ & \quad + \xi e\rho(\rho+2)b^\mu b^\nu R_{\mu\nu} + 2\xi e(\rho+1)b^\mu A^\nu R_{\mu\nu}\Big] \\ & -\frac{1}{4}eF_{\mu\nu}F^{\mu\nu} - eA_\mu J^\mu - eb_\mu J^\mu + eb^\mu\partial_\nu \Xi_\mu J^\nu\,,\end{aligned} \tag{49}$$

where again the Lorentz modes are relabeled as $A_\mu \equiv \mathcal{E}_\mu$, which obeys $b_\mu A^\mu = 0$, and the field strength is $F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu$. The gravitational excitations $h_{\mu\nu}$ obey the condition $h_{\mu\nu}b^\mu = 0$.

The form of this effective lagrangian reveals that only two of the four potential NG modes propagate. These are the transverse Lorentz NG modes. The longitudinal Lorentz and the diffeomorphism NG modes are auxiliary. In particular, the curvature terms do not provide kinetic terms for ρ. This is because, metric fluctuations in the form of a diffeomorphism excitation produce only a vanishing contribution to the curvature tensor at linear order.

In Riemann spacetime, the NG Lorentz modes again resemble the photon in an axial gauge. The interaction with the charged current J_μ also has the appropriate form. However, possible signals for testing the idea that the photon is due to Lorentz violation can be found. In particular, there are unconventional couplings of the curvature with A^μ, ρ, and b^μ. The curvature couplings $eA^\mu A^\nu R_{\mu\nu}$, are forbidden by gauge invariance in conventional Einstein-Maxwell electrodynamics, but they can appear here in a theory with Lorentz violation. The term $\xi e b^\mu b^\nu R_{\mu\nu}/2\kappa$ corresponds to an SME coefficient of the $s^{\mu\nu}$ type in the gravity sector of the SME. The remaining terms also represent Lorentz-violating couplings that are included in the SME. Any of these signals could serve to provide experimental evidence for the idea that the photon is an NG mode due to spontaneous Lorentz violation.

4.4 Riemann-Cartan Spacetime

In a Riemann-Cartan spacetime, the vierbein $e_\mu{}^a$ and the spin connection ω_μ^{ab} are independent degrees of freedom. As a result, the effects of spontaneous Lorentz breaking are very different from the cases of Minkowski and Riemann spacetime. In particular, it has been found that when the torsion is nonzero it is possible for a Higgs mechanism to occur [10]. This will be illustrated below in the context of the bumblebee model in Riemann-Cartan spacetime.

One immediate question concerning the possibility of a Higgs mechanism in a gravitational theory is whether the graviton acquires a mass or not. Indeed, even a small mass for the graviton can modify the predictions of general relativity leading to disagreement with experiment [38]. However, it was shown some time ago that a conventional Higgs mechanism cannot give rise to a mass for the graviton since the terms that are generated involve derivatives of the metric [11].

A generic Lagrangian for a theory with spontaneous Lorentz violation in Riemann-Cartan spacetime can be written as

$$\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_{\rm SSB}\,. \tag{50}$$

Here, we assume $\mathcal{L}_0$ contains only gravitational terms formed from the curvature and torsion and describes the unbroken theory, while $\mathcal{L}_{\rm SSB}$ induces spontaneous Lorentz violation. For a Higgs mechanism to occur involving the spin connection, $\mathcal{L}_0$ should describe massless propagating modes for the spin connection prior to the spontaneous breaking of Lorentz symmetry. The theory should also be free of ghosts. It turns out that these conditions severely restrict the possibilities for model building. The number of ghost-free theories with massive and massless propagating spin connection modes is limited [39, 40]. The number of propagating modes in these models depends on the presence of additional accidental

symmetries. The symmetry-breaking lagrangian $\mathcal{L}_{\rm SSB}$ typically breaks one or more of the accidental symmetries when the tensor field acquires a vev, which complicates the analysis of potential models.

In the bumblebee model in (31) the symmetry-breaking part of the lagrangian is

$$\mathcal{L}_{\rm SSB} = -\frac{1}{4}eB_{\mu\nu}B^{\mu\nu} - e\lambda(B_\mu B^\mu \pm b^2)\,. \tag{51}$$

In a Riemann-Cartan spacetime, the field strength $B_{\mu\nu}$ is defined in (32). In terms of the vierbein and spin connection, it becomes

$$B_{\mu\nu} = \left(e_\mu{}^b\omega_\nu{}^a{}_b - e_\nu{}^b\omega_\mu{}^a{}_b\right) b_a\,. \tag{52}$$

Note that this expression reduces back to (48) in the limits of Riemann and Minkowski spacetimes, where the spin connection is given by (47).

When $B_{\mu\nu}$ is squared, quadratic terms in $\omega_\mu{}^a{}_b$ appear in the Lagrangian, which perturbatively have the form

$$-\frac{1}{4}eB_{\mu\nu}B^{\mu\nu} \approx -\frac{1}{4}(\omega_{\mu\rho\nu} - \omega_{\nu\rho\mu})(\omega^{\mu\sigma\nu} - \omega^{\nu\sigma\mu})b^\rho b_\sigma\,. \tag{53}$$

It is these quadratic terms that suggest that a Higgs mechanism can occur involving the absorption of the NG modes by the spin connection. It should be noted that this is only possible in Riemann-Cartan spacetime with nonzero torsion, since otherwise (as in Riemann spacetime) the spin connection has no independent degrees of freedom.

In [10], a number of different models for the kinetic terms $\mathcal{L}_0$ were considered. As mentioned, the difficulty in building a viable model with a Higgs mechanism comes from finding a kinetic term describing propagating modes that are compatible with (53) as a mass term. If ghosts are permitted, this is straightforward. For example, with the choice

$$\mathcal{L}_0 = \frac{1}{4}R_{\lambda\kappa\mu\nu}R^{\lambda\kappa\mu\nu}\,. \tag{54}$$

all the fields $\omega_{\lambda\mu\nu}$ with $\lambda \neq 0$ propagate as massless modes. When this is combined with $\mathcal{L}_{\rm SSB}$, we find that among the propagating modes in the linearized theory there is a massive mode. Other examples can be studied as well and are aided by decomposing the fields $\omega_{\lambda\mu\nu}$ according to their spin-parity projections J^P in three-dimensional space. This reveals that the mass term consists of a physical 1^+ mode and a 1^- gauge mode. Models can be found in which $\mathcal{L}_0$ includes a massless 1^+ mode. However, typically the propagating massless modes involve combinations of J^P projections, which makes finding compatibility with $\mathcal{L}_{\rm SSB}$ all the more challenging.

In the end, a number of issues remain open for future investigation. Studies of the large variety of possible Lorentz-invariant lagrangians $\mathcal{L}_0$ can lead to new models in which the spin connection acquires a mass due to spontaneous Lorentz breaking. Different choices for $\mathcal{L}_{\rm SSB}$ can also be considered, including ones in which the spontaneous Lorentz violation involves one or more tensor fields. This

would certainly affect the dynamics of the NG modes as well. From a broader theoretical point of view, the incorporation of spontaneous Lorentz violation in theories with torsion opens up a new arena in the search for ghost-free models with propagating massive modes.

Certainly, there are implications for phenomenology in the context of Riemann-Cartan spacetime. The relevant mass scale in the Higgs mechanism is set by b^2. Even if this is on the order of the Planck mass, the existence of fields associated with Lorentz violation could have effects on cosmology, black holes, and gravitational radiation. Since all of the relevant terms in any of these models are included in the SME in Riemann-Cartan spacetime, a systematic approach would be to investigate possible new signals in that context.

5 Phenomenology

The minimal SME described in Sect. 3.1 has been used extensively in recent years by experimentalists and theorists to search for leading-order signals of Lorentz violation. To date, Planck-scale sensitivity has been attained to the dominant SME coefficients in a number of experiments involving different particle sectors. These include experiments with photons [29, 41–46], electrons [47–53], protons and neutrons [54–59], mesons [60,61], muons [62–64], neutrinos [8,17,65–67], and the Higgs [68]. It should be noted that despite the length of this list of experiments, a substantial portion of the SME coefficient space remains unexplored.

In the remaining sections, an overview of some of the recent tests of Lorentz and CPT symmetry in a Minkowski background will be given. In particular, since many of the sharpest test are performed in high-precision atomic and particle experiments involving photons and charged particles, much of the focus will be on the QED limit of the minimal SME. However, two other particle sectors are briefly described as well. These involve testing Lorentz and CPT symmetry with mesons and neutrinos.

5.1 Mesons

Experiments with mesons have long provided some of the sharpest tests of CPT. Since CPT and Lorentz symmetry are intertwined in field theory, these experiments also provide additional tests of Lorentz symmetry. Investigations in the context of the SME have found very high sensitivity to the CPT-odd a_μ coefficients in the SME.

The time evolution of a meson P^0 and its antimeson $\bar{P}^0$ is governed by a 2×2 effective hamiltonian Λ in a description based on the Schrödinger equation. Here, P represents one of the neutral mesons K, D, B_d, B_s. The hamiltonian can be written as [61, 69]

$$\Lambda = \frac{1}{2}\Delta\lambda \begin{pmatrix} U+\xi & VW^{-1} \\ VW & U-\xi \end{pmatrix} , \tag{55}$$

where the parameters U, V, W, and ξ are complex. The factor $\Delta\lambda/2$ ensures these parameters are dimensionless. Imposing conditions on the trace and determinant gives the relations $U \equiv \lambda/\Delta\lambda$ and $V \equiv \sqrt{1-\xi^2}$. The independent complex parameters $W = w\exp(i\omega)$ and $\xi = \mathrm{Re}[\xi]+i\mathrm{Im}[\xi]$ have four real components. One is physically unobservable. The argument ω changes under a phase redefinition of the P^0 wave function. The three others are physical. The two real numbers $\mathrm{Re}[\xi]$, $\mathrm{Im}[\xi]$ determine the amount of CPT violation, with CPT preserved if and only if both are zero.

The dominant CPT-violating contributions to the effective hamiltonian Λ can be calculated as expectation values of interaction terms in the SME. The result in terms of ξ is

$$\xi \sim \beta^\mu \Delta a_\mu \,, \tag{56}$$

where $\beta^\mu = \gamma(1, \boldsymbol{\beta})$ is the four-velocity of the P meson in the laboratory frame and the coefficients Δa_μ are combinations of SME coefficients.

The 4-velocity (and 4-momentum) dependence in (56) shows explicitly that CPT violation cannot be described with a constant complex parameter in quantum field theory [61]. Nonetheless, most experiments have fit their data to a constant value of ξ. Experiments in the kaon system [70], for example, have attained bounds of order 10^{-4} on the real and imaginary parts of ξ. More recently, however, analyses have been performed taking into account that in an experiment Δa_μ varies with the magnitude and direction of the momentum and with sidereal time as the Earth rotates. These experiments have attained sensitivities to Δa_μ on the order of 10^{-20} GeV in the kaon system and 10^{-15} GeV in the D system [60]. Additional bounds for the B_d and B_s systems can be obtained as well in future analyses.

5.2 Neutrinos

A general analysis in the context of the SME has searched for possible signals of Lorentz violation in neutrino physics [66]. Among other things, it looked at how free neutrinos with Dirac and Majorana couplings oscillate in the presence of Lorentz violation. Remarkably, a number of possible models exist in which Lorentz violation (with or without massive neutrinos) contributes to neutrino oscillations. One two-parameter model in particular, consisting of massless neutrinos, called the bicycle model, reproduces features in observed data (except for the LSND experiment). Indeed, a statistical analysis performed using data from Super-Kamiokande on atmospheric neutrinos finds that the fit based on the bicycle model is essentially as good (within a small marginal error) to the fit based on small mass differences [65]. Further investigations looking for sidereal time variations will be able to distinguish oscillations associated with Lorentz violation from those due to small mass differences.

5.3 QED Sector

Traditionally, many of the sharpest tests of Lorentz and CPT symmetry have been made with photons or in particle or atomic systems where the predominant

interactions are described by QED. This would include the original Michelson-Morley experiments and their modern-day versions [42–44]. The Lorentz tests known as Hughes-Drever experiments are atomic experiments in which two high-precision atomic clocks consisting of different atomic species are compared as the Earth rotates [54]. These provide exceptionally sharp tests of Lorentz symmetry. Similarly, some of the best CPT tests for leptons and baryons – involving direct comparisons of particles and antiparticles – are made by atomic physicists working with Penning traps [47, 48, 59].

In order to look for the leading order signals of Lorentz and CPT violation in these types of experiments, it is useful to work with a subset of the minimal SME lagrangian that is relevant to experiments in QED systems. The QED limit of the minimal SME can be written as

$$\mathcal{L}_{\rm QED} = \mathcal{L}_0 + \mathcal{L}_{\rm int}\,. \tag{57}$$

The lagrangian $\mathcal{L}_0$ contains the usual Lorentz-invariant terms in QED describing photons, massive charged fermions, and their conventional couplings, while $\mathcal{L}_{\rm int}$ contains the Lorentz-violating interactions. Since the minimal SME in flat spacetime is restricted to the remormalizable and gauge-invariant terms in the full SME, the QED sector interactions in $\mathcal{L}_{\rm int}$ have a finite number of terms. For the case of photons and a single fermion species ψ the Lorentz-violating terms are given by [71]

$$\begin{aligned}\mathcal{L}_{\rm int} = &-a_\mu\bar\psi\gamma^\mu\psi - b_\mu\bar\psi\gamma_5\gamma^\mu\psi + ic_{\mu\nu}\bar\psi\gamma^\mu D^\nu\psi \\ &+id_{\mu\nu}\bar\psi\gamma_5\gamma^\mu D^\nu\psi - \frac{1}{2}H_{\mu\nu}\bar\psi\sigma^{\mu\nu}\psi \\ &-\frac{1}{4}(k_F)_{\kappa\lambda\mu\nu}F^{\kappa\lambda}F^{\mu\nu} + \frac{1}{2}(k_{AF})^\kappa\epsilon_{\kappa\lambda\mu\nu}A^\lambda F^{\mu\nu}\,.\end{aligned} \tag{58}$$

Here, $iD_\mu \equiv i\partial_\mu - qA_\mu$. The terms with coefficients a_μ, b_μ and $(k_{AF})_\mu$ are odd under CPT, while those with $H_{\mu\nu}$, $c_{\mu\nu}$, $d_{\mu\nu}$, and $(k_F)_{\kappa\lambda\mu\nu}$ preserve CPT. All seven terms break Lorentz symmetry. In general, superscript labels will be added to these parameters to denote the particle species.

This Lagrangian emerges naturally from the minimal SME sector for charged leptons, following the usual assumptions of electroweak symmetry breaking and mass generation. Lagrangian terms of the same form are expected to describe protons and neutrons in QED systems as well, but where the SME coefficients represent composites stemming from quark and gluon interactions. It is certainly the case that QED and its relativistic quantum-mechanical limits describe proton and neutron electromagnetic interactions in atoms in excellent agreement with experiments. Defining terms involving composite SME parameters for protons and neutrons is therefore a reasonable extension of the theory. The QED extension of the SME treats protons and neutrons as the basic constituents of the theory. The lagrangian $\mathcal{L}_{\rm int}$ then contains the most general set of Lorentz-violating interactions in this context.

Since the corrections due to Lorentz violation at low energy are known to be small, it is sufficient in many situations to work in the context of relativistic

quantum mechanics using perturbation theory. To do so, a Hamiltonian is needed such that

$$i\partial_0\chi = \hat{H}\chi\,, \tag{59}$$

where $\hat{H} = \hat{H}_0 + \hat{H}_{\rm pert}$. The perturbative hamiltonian $\hat{H}_{\rm pert}$ associated with Lorentz violation can be generated using a Foldy-Wouthuysen approach and by making appropriate field redefinitions [49, 57]. The result for a massive fermion particle is

$$\begin{aligned}\hat{H}_{\rm pert} = {} & a_\mu\gamma^0\gamma^\mu - b_\mu\gamma_5\gamma^0\gamma^\mu - c_{00}m\gamma^0 - i(c_{0j}+c_{j0})D^j \\ & + i(c_{00}D_j - c_{jk}D^k)\gamma^0\gamma^j - d_{j0}m\gamma_5\gamma^j + i(d_{0j}+d_{j0})D^j\gamma_5 \\ & + i(d_{00}D_j - d_{jk}D^k)\gamma^0\gamma_5\gamma^j + \frac{1}{2}H_{\mu\nu}\gamma^0\sigma^{\mu\nu}\,. \end{aligned} \tag{60}$$

Here, the letters j, k, l, etc. represent the three spatial directions in a laboratory frame. The $j = 3$ (or z direction) is usually chosen as the quantization axis. The corresponding hamiltonian for the antiparticle can be obtained using charge conjugation.

The SME coefficients are expected to be fixed with respect to a nonrotating coordinate frame. As a result, the SME coefficients b_0, b_j, etc. would change as the Earth moves. In order to give measured bounds in a consistent manner, a nonrotating frame is chosen. Often, this is chosen as a sun-centered frame using celestial equatorial coordinates. These are denoted using upper-case letters T,X,Y,Z. Typically, experiments sensitive to sidereal time variations are sensitive to a combination of parameters, which are denoted using tildes. For example, the b_μ tilde coefficients with $\mu = j$ are defined as

$$\tilde{b}^e_j \equiv b^e_j - md^e_{j0} - \frac{1}{2}\varepsilon_{jkl}H^e_{kl}\,, \tag{61}$$

These combinations are projected onto the nonrotating frame, where the components with respect to the celestial equatorial coordinate frame are b^e_X, b^e_Y, b^e_Z, etc. The relation between the laboratory and nonrotating components is

$$\begin{aligned}\tilde{b}^e_1 &= \tilde{b}^e_X\cos\chi\cos(\Omega t) + \tilde{b}^e_Y\cos\chi\sin(\Omega t) - \tilde{b}^e_Z\sin\chi\,, \\ \tilde{b}^e_2 &= -\tilde{b}^e_X\sin(\Omega t) + \tilde{b}^e_Y\cos(\Omega t)\,, \\ \tilde{b}^e_3 &= \tilde{b}^e_X\sin\chi\cos(\Omega t) + \tilde{b}^e_Y\sin\chi\sin(\Omega t) + \tilde{b}^e_Z\cos\chi\,. \end{aligned} \tag{62}$$

The angle χ is between the $j = 3$ lab axis and the direction of the Earth's rotation axis, which points along Z. The angular frequency $\Omega \simeq 2\pi/(23\text{h}\,56\text{m})$ is that corresponding to a sidereal day.

6 Tests in QED

Before examining individual tests of Lorentz symmetry in QED systems, it is useful to examine some of the more general results that have emerged from these

investigations. One general feature is that sensitivity to Lorentz and CPT violation in these experiments stems primarily from their ability to detect very small anomalous energy shifts. While many of the experiments were originally designed to measure specific quantities, such as charge-to-mass ratios of particles and antiparticles or differences in g factors, it is now recognized that these experiments are most effective as Lorentz and CPT tests when all of the energy levels in the system are investigated for possible anomalous shifts. As a result of this, a number of new signatures of Lorentz and CPT violation have been discovered in recent years that were overlooked previously.

A second general feature concerns how these atomic experiments are typically divided into two groups. The first (Lorentz tests) looks for sidereal time variations in the energy levels of a particle or atom. The second (CPT tests) looks for a difference in the energy levels between a particle (or atom) and its antiparticle (or antiatom). What has been found is that the sensitivity to Lorentz and CPT violation in these two classes of experiments is not distinct. Experiments traditionally viewed as Lorentz tests are also sensitive to CPT symmetry and vice versa. Nonetheless, it is important to keep in mind that that there are differences as well. For example, the CPT experiments comparing matter and antimatter are directly sensitive to CPT-violating parameters, such as b_μ, whereas Lorentz tests are sensitive to combinations of CPT-preserving and CPT-violating parameters, which are denoted using a tilde. Ultimately, both clases of experiments are important and should be viewed as complementary.

It has become common practice to express sensitivities to Lorentz and CPT violation in terms of the SME coefficients. This provides a straightforward approach that allows comparisons across different types of experiments. Since each different particle sector in the QED extension has an independent set of Lorentz-violating SME coefficients, these are distinguished using superscript labels. A thorough investigation of Lorentz and CPT violation necessarily requires looking at as many different particle sectors as possible.

6.1 Photons

The lagrangian describing a freely propagating photon in the presence of Lorentz violation is given by [45]

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{4}(k_F)_{\kappa\lambda\mu\nu}F^{\kappa\lambda}F^{\mu\nu} + \frac{1}{2}(k_{AF})^{\kappa}\epsilon_{\kappa\lambda\mu\nu}A^{\lambda}F^{\mu\nu}\,, \qquad (63)$$

where the field strength $F_{\mu\nu}$ is defined by $F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu$.

The coefficient k_{AF}, which is odd under CPT, has been investigated extensively both theoretically and experimentally [41, 45]. Theoretically, it is found that this term leads to negative-energy contributions and is a potential source of instability in the theory. One solution is to set k_{AF} to zero, which has been shown to be consistent with radiative corrections in the SME. However, stringent experimental constraints also exist consistent with $k_{AF} \approx 0$. These result from studying the polarization of radiation from distant radio galaxies. In what follows, we will therefore ignore the effects of the k_{AF} terms.

The terms with coefficients k_F, which is even under CPT, have been investigated more recently [45]. These terms provide positive-energy contributions. There are 19 independent components in the k_F coefficients. It is useful to rewrite them in terms of a new set, $\tilde{\kappa}_{e+}$, $\tilde{\kappa}_{e-}$, $\tilde{\kappa}_{o+}$, $\tilde{\kappa}_{o-}$, and $\tilde{\kappa}_{\rm tr}$. Here, $\tilde{\kappa}_{e+}$, $\tilde{\kappa}_{e-}$, and $\tilde{\kappa}_{o-}$ are 3×3 traceless symmetric matrices (with 5 independent components each), while $\tilde{\kappa}_{o+}$ is a 3×3 antisymmetric matrix (with 3 independent components), and the remaining coefficient $\tilde{\kappa}_{\rm tr}$ is the only rotationally invariant component.

The lagrangian can be written in terms of the new set and the usual electric and magnetic fields $\mathbf{E}$ and $\mathbf{B}$ as follows:

$$\mathcal{L} = \frac{1}{2}[(1+\tilde{\kappa}_{\rm tr})\mathbf{E}^2 - (1-\tilde{\kappa}_{\rm tr})\mathbf{B}^2] + \frac{1}{2}\mathbf{E}\cdot(\tilde{\kappa}_{e+}+\tilde{\kappa}_{e-})\cdot\mathbf{E} - \frac{1}{2}\mathbf{B}\cdot(\tilde{\kappa}_{e+}-\tilde{\kappa}_{e-})\cdot\mathbf{B} + \mathbf{E}\cdot(\tilde{\kappa}_{o+}+\tilde{\kappa}_{o-})\cdot\mathbf{B}\,. \tag{64}$$

This lagrangian gives rise to modifications of Maxwell's equations, which have been explored in recent astrophysical and laboratory experiments. Ten of the coefficients, $\tilde{\kappa}_{e+}$ and $\tilde{\kappa}_{o-}$, lead to birefrigence of light. Bounds on these parameters of order 2×10^{-32} have been obtained from spectropolarimetry of light from distant galaxies [45]. The nine coefficients, $\tilde{\kappa}_{\rm tr}$, $\tilde{\kappa}_{e-}$, and $\tilde{\kappa}_{o+}$, have been bounded in a series of recent laboratory photon experiments. Seven of the eight $\tilde{\kappa}_{e-}$ and $\tilde{\kappa}_{o+}$ coefficients, have been bounded in experiments using optical and microwave cavities. Sensitivities on the order of $\tilde{\kappa}_{o+} \lesssim 10^{-11}$ and $\tilde{\kappa}_{e-} \lesssim 10^{-15}$ have been attained [42]. The trace coefficient has been estimated to have an upper bound of $\tilde{\kappa}_{\rm tr} \lesssim 10^{-4}$ from Ives-Stilwell experiments [43]. The remaining $\tilde{\kappa}_{e-}$ coefficient has recently been bounded at the level of 10^{-14} using a rotating apparatus [44].

6.2 Penning Traps

There are primarily two leading-order signals of Lorentz and CPT violation that can be searched for in experiments in Penning traps [49]. One is a traditional CPT test, comparing particles and antiparticles, while the other is a Lorentz test that looks for sidereal time variations. Both types of signals have been investigated in recent years in experiments with electrons and positrons. The experiments involve making high-precision measurements of the anomaly frequency ω_a and the cyclotron frequency ω_c of the trapped electrons and/or positrons.

The first test was a reanalysis was performed by Dehmelt's group using existing data for electrons and positrons in a Penning trap [47]. The idea was to look for an instantaneous difference in the anomaly frequencies of electrons and positrons, which can be nonzero when CPT and Lorentz symmetry are broken. Dehmelt's original measurements of $g-2$ did not involve looking for possible instantaneous variations in ω_a. Instead, the ratio ω_a/ω_c was computed using averaged values. However, Lorentz-violating corrections to the anomaly frequency

ω_a can occur even if the g factor remains unchanged. An alternative analysis therefore looks for an instantaneous difference in the electron and positron anomaly frequencies. The new bound found by Dehmelt's group can be expressed in terms of the parameter b_3^e, which is the component of b_μ^e along the quantization axis in the laboratory frame. The bound they obtained is $|b_3^e| \lesssim 3 \times 10^{-25}$ GeV.

The second signal for Lorentz and CPT violation in the electron sector involves measurements of the electron alone [48]. Here, the idea is that the Lorentz and CPT-violating interactions depend on the orientation of the quantization axis in the laboratory frame, which changes as the Earth turns on its axis. As a result, both the cyclotron and anomaly frequencies have small corrections which cause them to exhibit sidereal time variations. Such a signal can be measured using just electrons, which eliminates the need for comparison with positrons. The bounds in this case are given with respect to a nonrotating coordinate system such as celestial equatorial coordinates. The interactions involve a combination of laboratory-frame components that couple to the electron spin. The combination is denoted as $\tilde{b}_3^e \equiv b_3^e - m d_{30}^e - H_{12}^e$. The bound can be expressed in terms of components X, Y, Z in the nonrotating frame. It is given as $|\tilde{b}_J^e| \lesssim 5 \times 10^{-25}$GeV for $J = X, Y$.

Although no $g-2$ experiments have been made for protons or antiprotons, there have been recent bounds obtained on Lorentz violation in comparisons of cyclotron frequencies of antiprotons and H^- ions confined in a Penning trap [59]. In this case the sensitivity is to the dimensionless parameters $c_{\mu\nu}^p$. Future experiments with protons and antiprotons will be able to provide tests that are sensitive to b_μ^p.

6.3 Clock-Comparison Experiments

The classic Hughes-Drever experiments are atomic clock-comparison tests of Lorentz invariance [54,57]. There have been a number of different types of these experiments performed over the years, with steady improvements in their sensitivity. They involve making high-precision comparisons of atomic clock signals as the Earth rotates. The clock frequencies are typically hyperfine or Zeeman transitions. Many of the sharpest Lorentz bounds for the proton, neutron, and electron stem from atomic clock-comparison experiments. For example, Bear et al. in [54] used a two-species noble-gas maser to test for Lorentz and CPT violation in the neutron sector. They obtain a bound $|\tilde{b}_J^n| \lesssim 10^{-31}$ GeV for $J = X, Y$, which is currently the best bound for the neutron sector.

It should also be pointed out that certain assumptions about the nuclear configurations must be made to obtain bounds in clock-comparison experiments. For this reason, these bounds should be viewed as good to within about an order of magnitude. To obtain cleaner bounds it is necessary to consider simpler atoms or to perform more sophisticated nuclear modeling.

Note as well that these Earth-based laboratory experiments are not sensitive to Lorentz-violation coefficients along the $J = Z$ direction parallel to Earth's rotation axis. They also neglect the velocity effects due to Earth's motion around

the sun, which would lead to bounds on the timelike components along $J = T$. These limitations can be overcome by performing experiments in space or by using a rotation platform. The earth's motion can also be taken into account. A recent boosted-frame analysis of the dual noble-gas maser experiment has yielded bounds on the order of 10^{-27} GeV on many boost-dependent SME coefficients for the neutron that were previously unbounded [56].

6.4 Experiments in Space

Clock-comparison experiments performed in space would have several advantages over traditional ground-based experiments [58]. For example, a clock-comparison experiment conducted aboard the International Space Station (ISS) would be in a laboratory frame that is both rotating and boosted. It would therefore immediately gain sensitivity to both the Z and timelike directions. This would more than triple the number of Lorentz-violation parameters that are accessible in a clock-comparison experiment. Another advantage of an experiment aboard the ISS is that the time needed to acquire data would be greatly reduced (by approximately a factor of 16). In addition, new types of signals would emerge that have no analogue in traditional Earth-based experiments. The combination of these advantages should result in substantially improved limits on Lorentz and CPT violation. Unfortunately, the USA has canceled its missions aimed at testing fundamental physics aboard the ISS. However, there is still a European mission planned for the ISS which will compare atomic clocks and H masers. Therefore, the opportunity to perform these new Lorentz and CPT tests is still a possibility.

6.5 Hydrogen and Antihydrogen

Hydrogen atoms have the simplest nuclear structure, and antihydrogen is the simplest antiatom. These atoms (or antiatoms) therefore provide opportunities for conducting especially clean Lorentz and CPT tests involving protons and electrons.

There are three experiments underway at CERN that can perform high-precision Lorentz and CPT tests in antihydrogen [12]. Two of the experiments (ATRAP and ATHENA) intend to make high-precision spectroscopic measurements of the 1S-2S transitions in hydrogen and antihydrogen. These are forbidden (two-photon) transitions that have a relative linewidth of approximately 10^{-15}. The ultimate goal is to measure the line center of this transition to a part in 10^3 yielding a frequency comparison between hydrogen and antihydrogen at a level of 10^{-18}. An analysis of the 1S-2S transition in the context of the SME shows that the magnetic field plays an important role in the attainable sensitivity to Lorentz and CPT violation [50]. For instance, in free hydrogen in the absence of a magnetic field, the 1S and 2S levels are shifted by equal amounts at leading order. As a result, in free H or $\bar{\mathrm{H}}$ there are no leading-order corrections to the 1S-2S transition frequency. In a magnetic trap, however, there are fields that

can mix the spin states in the four different hyperfine levels. Since the Lorentz-violating interactions depend on the spin orientation, there will be leading-order sensitivity to Lorentz and CPT violation in comparisons of 1S-2S transitions in trapped hydrogen and antihydrogen. At the same time, however, these transitions are field-dependent, which creates additional experimental challenges that would need to be overcome.

An alternative to 1S-2S transitions is to consider the sensitivity to Lorentz violation in ground-state Zeeman hyperfine transitions. It is found that there are leading-order corrections in these levels in both hydrogen and antihydrogen [50]. The ASACUSA group at CERN is planning to measure the Zeeman hyperfine transitions in antihydrogen. Such measurements will provide a direct CPT test.

Experiments with hydrogen alone have been performed using a maser [55]. They attain exceptionally sharp sensitivity to Lorentz and CPT violation in the electron and proton sectors of the SME. These experiments use a double-resonance technique that does not depend on there being a field-independent point for the transition. The sensitivity for the proton attained in these experiments is $|\tilde{b}_J^p| \lesssim 10^{-27}$ GeV. Due to the simplicity of hydrogen, this is an extremely clean bound and is currently the most stringent test of Lorentz and CPT violation for the proton.

6.6 Muon Experiments

Experiments with muons involve second-generation leptons and provide tests of CPT and Lorentz symmetry that are independent of the tests involving electrons. There are several different types of experiments with muons that have recently been conducted, including muonium experiments [62] and $g-2$ experiments with muons at Brookhaven [63]. In muonium, experiments measuring the frequencies of ground-state Zeeman hyperfine transitions in a strong magnetic field have the greatest sensitivity to Lorentz and CPT violation. A recent analysis has searched for sidereal time variations in these transitions. A bound at the level of $|\tilde{b}_J^\mu| \leq 2 \times 10^{-23}$ GeV has been obtained [62]. In relativistic $g-2$ experiments using positive muons with "magic" boost parameter $\delta = 29.3$, bounds on Lorentz-violation parameters are possible at a level of 10^{-25} GeV. However, the analysis of these experiments is still underway at Brookhaven.

6.7 Spin Polarized Torsion Pendulum

Experiments using spin polarized torsion pendula have been conducted at the University of Washington and in Taiwan. These experiments currently provide the sharpest bounds on Lorentz and CPT symmetry in the electron sector [52]. These experiments are able to achieve very high sensitivity to Lorentz violation because the torsion pendula have a huge number of aligned electron spins but a negligible magnetic field.

The pendulum at the University of Washington is built out of a stack of toroidal magnets, which in one version of the experiment achieved a net electron

spin $S \simeq 8 \times 10^{22}$. The apparatus is suspended on a rotating turntable and the time variations of the twisting pendulum are measured. An analysis of this system shows that in addition to a signal having the period of the rotating turntable, the effects due to Lorentz and CPT violation also cause additional time variations with a sidereal period caused by the rotation of the Earth. The group at the University of Washington has analyzed data taken in 1998 and find that thay have sensitivity to the electron coefficients at the levels of $|\tilde{b}_J^e| \lesssim 10^{-29}$ GeV for $J = X, Y$ and $|\tilde{b}_Z^e| \lesssim 10^{-28}$ GeV. More recently, a new pendulum has been built, and it is expected that 20-fold improved sensitivities will be attained [72].

The Taiwan experiment also uses a rotating torsion pendulum, which is made of a ferrimagnetic material. This group achieved a net polarization of $S \simeq 8.95 \times 10^{22}$ electrons in their pendulum. The bounds they obtain for the electron are at the levels of $|\tilde{b}_J^e| \lesssim 3.1 \times 10^{-29}$ GeV for $J = X, Y$ and $|\tilde{b}_Z^e| \lesssim 7.1 \times 10^{-28}$ GeV.

7 Conclusions

This overview describes the development and use of the SME as the theoretical framework describing Lorentz violation in the context of field theory. The philosophy of the SME is that any interactions that are observer invariant and involve known fields at low energy are included in the theory. As an incremental first step, the minimal SME (and its QED limit) can be constructed. This theory maintains gauge invariance and power-counting renormalizability. It is the suitable framework for investigating leading-order signals of Lorentz violation.

In addition to constructing the SME, we have examined the special case of spontaneous Lorentz breaking. In particular, the question of what the fate of the Nambu-Goldstone modes is when Lorentz symmetry is spontaneously broken has been addressed. We have demonstrated that spontaneous particle Lorentz violation is accompanied by spontaneous particle diffeomorphism violation and vice versa, and that up to 10 NG modes can appear. These modes can comprise 10 of the 16 modes of the vierbein that in a Lorentz-invariant theory are gauge degrees of freedom. The fate of the NG modes is found to depend also on the spacetime geometry and on the behavior of the tensor vev inducing spontaneous Lorentz violation. These results have been illustrated using a bumblebee model. In Minkowski and Riemann spacetimes, it is found that the NG modes propagate like the photon in an axial gauge. In Riemann-Cartan spacetimes, the interesting possibility exists that the spin connection could absorb the propagating NG modes in a gravitational version of the Higgs mechanism. This unique feature of gravity theories with torsion may offer another phenomenologically viable route for constructing realistic models with spontaneous Lorentz violation.

Phenomenology has been investigated using the minimal SME. Experiments in QED systems continue to provide many of the sharpest tests of Lorentz and CPT symmetry. In recent years, a number of new astrophysical and laboratory tests have been performed that have lead to substantially improved sensitivities

Table 1. Summary of leading-order bounds for the parameter $\tilde{b}_J$.

Expt	Sector	Params ($J = X, Y$)	Bound (GeV)
Penning Trap	electron	$\tilde{b}_J^e$	5×10^{-25}
Hg-Cs clock comparison	electron	$\tilde{b}_J^e$	$\sim 10^{-27}$
	proton	$\tilde{b}_J^p$	$\sim 10^{-27}$
	neutron	$\tilde{b}_J^n$	$\sim 10^{-30}$
He-Xe dual maser	neutron	$\tilde{b}_J^n$	$\sim 10^{-31}$
H maser	electron	$\tilde{b}_J^e$	10^{-27}
	proton	$\tilde{b}_J^p$	10^{-27}
Muonium	muon	$\tilde{b}_J^\mu$	2×10^{-23}
Spin Pendulum	electron	$\tilde{b}_J^e$	10^{-29}
		$\tilde{b}_Z^e$	10^{-28}

for the photon. Similarly, atomic experimentalists continue to find ways of improving the sensitivity to Lorentz violation in many of the matter sectors of the SME. For comparison across different atomic experiments a summary of recent bounds on the $\tilde{b}_J$ coefficients in the minimal SME is given in Table 1. These bounds are within the range of sensitivity associated with suppression factors arising from the Planck scale. A more complete table would list all of the coefficients in the minimal SME. Note that many SME coefficients have still not been measured. Future experiments, in particular those performed in boosted frames, are likely to provide sensitivity to many of these currently unmeasured SME coefficients. In addition, the overall sensitivity of these experiments is expected to improve over the coming years.

References

1. A. Einstein, *The Principle of Relativity*, (Dover, New York, 1952).
2. R. Utiyama, Phys. Rev. **101**, 1597 (1956); T.W.B. Kibble, J. Math. Phys. **2**, 212 (1961).
3. For reviews of gravitation in Riemann-Cartan spacetimes see, for example, F.W. Hehl et al., Rev. Mod. Phys. **48**, 393 (1976); I.L. Shapiro, Phys. Rep. **357**, 113 (2002).
4. V.A. Kostelecký and S. Samuel, Phys. Rev. D **39**, 683 (1989); V.A. Kostelecký and R. Potting, Nucl. Phys. B **359**, 545 (1991).
5. V.A. Kostelecký and S. Samuel, Phys. Rev. Lett. **66**, 1811 (1991); V.A. Kostelecký and R. Potting, Phys. Lett. B **381**, 89 (1996); Phys. Rev. D **63**, 046007 (2001); V.A. Kostelecký, M. Perry, and R. Potting, Phys. Rev. Lett. **84**, 4541 (2000).

6. See, for example, R. Gambini and J. Pullin, Phys. Rev. D **59**, 124021 (1999); J. Alfaro, H.A. Morales-Técotl, and L.F. Urrutia, Phys. Rev. D **66**, 124006 (2002); D. Sudarsky, L. Urrutia, and H. Vucetich, Phys. Rev. Lett. **89**, 231301 (2002); Phys. Rev. D **68**, 024010 (2003); G. Amelino-Camelia, Mod. Phys. Lett. A **17**, 899 (2002); Y.J. Ng, Mod. Phys. Lett. **A18**, 1073 (2003); R. Myers and M. Pospelov, Phys. Rev. Lett. **90**, 211601 (2003); N.E. Mavromatos, Nucl. Instrum. Meth. B **214**, 1 (2004).
7. V.A. Kostelecký and R. Potting, Phys. Rev. D **51**, 3923 (1995).
8. D. Colladay and V.A. Kostelecký, Phys. Rev. D **55**, 6760 (1997); Phys. Rev. D **58**, 116002 (1998).
9. V.A. Kostelecký, Phys. Rev. D **69**, 105009 (2004).
10. R. Bluhm and V.A. Kostelecký, Phys. Rev. D **71**, 065008 (2005).
11. V.A. Kostelecký and S. Samuel, Phys. Rev. Lett. **63**, 224 (1989); Phys. Rev. D **40**, 1886 (1989).
12. For recent reviews of various experimental and theoretical approaches to Lorentz and CPT violation see, for example, V.A. Kostelecký, ed., *CPT and Lorentz Symmetry III* (World Scientific, Singapore, 2005) and earlier volumes in this series: *CPT and Lorentz Symmetry II*, World Scientific, Singapore, 2002; *CPT and Lorentz Symmetry*, World Scientific, Singapore, 1999.
13. C. Lämmerzahl, C.W.F. Everitt, F.W. Hehl, eds., *Gyros, Clocks, Interferometers: Testing Relativistic Gravity in Space* (Springer, Berlin, 2001)
14. Examples of some recent review articles are C. Lämmerzahl, A. Macias, and H. Müller, gr-qc/0501048; G. Amelino-Camelia, C. Lämmerzahl, A. Macias, and H. Müller, gr-qc/0501053; D. Mattingly, gr-qc/0502097; C. W. Will, gr-qc/0504085; gr-qc/0504086; T. Jacobson, S. Liberati, and D. Mattingly, astro-ph/0505267.
15. A. Connes, M. Douglas, and A. Schwartz, J. High Energy Phys. **02**, 003 (1998).
16. See, for example, I. Mocioiu, M. Pospelov, and R. Roiban, Phys. Lett. B **489**, 390 (2000); S.M. Carroll et al., Phys. Rev. Lett. **87**, 141601 (2001); Z. Guralnik, R. Jackiw, S.Y. Pi, and A.P. Polychronakos, Phys. Lett. B **517**, 450 (2001); C.E. Carlson, C.D. Carone, and R.F. Lebed, Phys. Lett. B **518**, 201 (2001); A. Anisimov, T. Banks, M. Dine, and M. Graesser, Phys. Rev. D **65**, 085032 (2002).
17. S. Coleman and S.L. Glashow, Phys. Rev. D **59**, 116008 (1999).
18. R.C. Myers and M. Pospelov, Phys. Rev. Lett. **90**, 211601 (2003).
19. T. Jacobson and D. Mattingly, Phys. Rev. D **70**, 024003 (2004).
20. A.P. Lightman and D.L. Lee, Phys. Rev. D **8**, 364 (1973).
21. H.P. Robertson, Rev. Mod. Phys. **21**, 378 (1949); R. Mansouri and R.U. Sexl, Gen. Rel. Grav. **8**, 497 (1977).
22. C.N. Will, *Theory and experimentation in Gravitational Physics* (Cambridge University Press, Cambridge, England, 1993).
23. J. Schwinger, Phys. Rev. **82** (1951) 914; J.S. Bell, Birmingham University thesis (1954); Proc. Roy. Soc. (London) **A 231** (1955) 479; G. Lüders, Det. Kong. Danske Videnskabernes Selskab Mat.fysiske Meddelelser **28**, No. 5 (1954); Ann. Phys. (N.Y.) **2** (1957) 1; W. Pauli, in W. Pauli, ed., *Neils Bohr and the Development of Physics*, McGraw-Hill, New York, 1955, p. 30.
24. O.W. Greenberg, Phys. Rev. Lett. **89**, 231602 (2002); Phys. Lett. B **567**, 179 (2003).
25. V.A. Kostelecký and R. Lehnert, Phys. Rev. D **63**, 065008 (2001).
26. D. Colladay and P. McDonald, J. Math. Phys. **43**, 3554 (2002).
27. M.S. Berger and V.A. Kostelecký, Phys. Rev. D **65**, 091701(R) (2002).
28. V.A. Kostelecký, C.D. Lane, and A.G.M. Pickering, Phys. Rev. D **65**, 056006 (2002);

29. R. Jackiw and V.A. Kostelecký, Phys. Rev. Lett. **82**, 3572 (1999); M. Pérez-Victoria, JHEP **0104**, 032 (2001).
30. V.A. Kostelecký, R. Lehnert, and M. Perry, Phys. Rev. D **68**, 123511 (2003).
31. Y. Nambu, Phys. Rev. Lett. **4**, 380 (1960); J. Goldstone, Nuov. Cim. **19**, 154 (1961); J. Goldstone, A. Salam, and S. Weinberg, Phys. Rev. **127**, 965 (1962).
32. F. Englert and R. Brout, Phys. Rev. Lett. **13**, 321 (1964); P.W. Higgs, Phys. Rev. Lett. **13**, 508 (1964); G.S. Guralnik, C.R. Hagen, and T.W.B. Kibble, Phys. Rev. Lett. **13**, 585 (1964).
33. See, for example, H.B. Nielsen and S. Chadha, Nucl. Phys. B **105**, 445 (1976); I. Low and A.V. Manohar, Phys. Rev. Lett. **88**, 101602 (2002); Y. Nambu, in *CPT and Lorentz Symmetry III*, [12].
34. T. Jacobson and D. Mattingly, Phys. Rev. D **64**, 024028 (2001); P. Kraus and E.T. Tomboulis, Phys. Rev. D **66**, 045015 (2002); J.W. Moffat, Intl. J. Mod. Phys. D **2**, 351 (1993); Found. Phys. **23** 411 (1993); Intl. J. Mod. Phys. D **12**, 1279 (2003); C. Eling and T. Jacobson, Phys. Rev. D **69**, 064005 (2004); A. Jenkins, Phys. Rev. D **69**, 105007 (2004); S.M. Carroll and E.A. Lim, Phys. Rev. D **70**, 123525 (2004); E.A. Lim, Phys. Rev. D **71**, 063504 (2005); B.M. Gripaios, JHEP **0410**, 069 (2004); J.L. Chkareuli, C.D. Froggatt, R.N. Mohapatra, and H.B. Nielsen, hep-th/0412225; M.L. Graesser, A. Jenkins, M.B. Wise, Phys. Lett. B B **613**, 5 (2005); O. Bertolami and J. Paramos, hep-th/0504215.
35. N. Arkani-Hamed, H.-C. Cheng, M. Luty, and J. Thaler, JHEP **0405**, 074 (2004).
36. P.A.M. Dirac, Proc. R. Soc. Lon. **A209**, 291, (1951); W. Heisenberg, Rev. Mod. Phys. **29**, 269 (1957); P.G.O. Freund, Acta Phys. Austriaca **14**, 445 (1961); J.D. Bjorken, Ann. Phys. **24**, 174 (1963).
37. Y. Nambu, Prog. Theor. Phys. Suppl. Extra 190 (1968).
38. H. van Dam and M. Veltman, Nucl. Phys. B **22**, 397 (1970); V.I. Zakharov, JEPT Lett. **12**, 312 (1970). A recent discussion of the discontinuity in a non-Minkowski background is F.A. Dilkes, M.J. Duff, J.T. Liu, and H. Sati, Phys. Rev. Lett. **87**, 041301 (2001).
39. E. Sezgin and P. van Nieuwenhuizen, Phys. Rev. D **21**, 3269 (1980).
40. K. Fukuma, Prog. Theor. Phys. **107**, 191 (2002).
41. S.M. Carroll, G.B. Field, and R. Jackiw, Phys. Rev. D **41**, 1231 (1990); M.P. Haugan and T.F. Kauffmann, Phys. Rev. D **52**, 3168 (1995).
42. J. Lipa et al., Phys. Rev. Lett. **90**, 060403 (2003); H. Müller et al., Phys. Rev. Lett. **91**, 020401 (2003); P. Wolf et al., Gen. Rel. Grav. **36**, 2351 (2004); Phys. Rev. D **70**, 051902 (2004).
43. M.E. Tobar et al., Phys. Rev. D **71**, 025004 (2005).
44. P. Antonini et al., gr-qc/0504109.
45. V.A. Kostelecký and M. Mewes, Phys. Rev. Lett. **87**, 251304 (2001); Phys. Rev. D **66**, 056005 (2002).
46. C. Adam and F.R. Klinkhamer, Nucl. Phys. B **657**, 214 (2003); H. Müller et al., Phys. Rev. D **67**, 056006 (2003); T. Jacobson, S. Liberati, and D. Mattingly, Phys. Rev. D **67**, 124011 (2003); V.A. Kostelecký and A.G.M. Pickering, Phys. Rev. Lett. **91**, 031801 (2003); R. Lehnert, Phys. Rev. D **68**, 085003 (2003); G.M. Shore, Contemp. Phys. **44**, 503 2003; B. Altschul, Phys. Rev. D **69**, 125009 (2004); Phys. Rev. D **70**, 101701 (2004); hep-th/0402036; T. Jacobson, S. Liberati, D. Mattingly, and F. Stecker, Phys. Rev. Lett. **93**, 021101 (2004); R. Lehnert and R. Potting, Phys. Rev. Lett. **93**, 110402 (2004); hep-ph/0408285; F.R. Klinkhamer and C. Rupp, Phys. Rev. D **70**, 045020 (2004); Q. Bailey and V.A. Kostelecký, Phys. Rev. D **70**, 076006 (2004); C. Lämmerzahl, A. Macias, and H. Müller, Phys. Rev.

D, in press; C. Lämmerzahl and F.W. Hehl, Phys. Rev. D **70**, 105022 (2004); H. Belich, T. Costa-Soares, M.M. Ferreira, and J.A. Helayel-Neto, hep-th/0411151; C. Lane, hep-ph/0505130.
47. H. Dehmelt et al., Phys. Rev. Lett. **83**, 4694 (1999).
48. R. Mittleman et al., Phys. Rev. Lett. **83**, 2116 (1999).
49. R. Bluhm, V.A. Kostelecký, and N. Russell, Phys. Rev. Lett. **79**, 1432 (1997); Phys. Rev. D **57**, 3932 (1998).
50. R. Bluhm, V.A. Kostelecký, and N. Russell, Phys. Rev. Lett. **82**, 2254 (1999).
51. D. Colladay and V.A. Kostelecký, Phys. Lett. B **511**, 209 (2001); B. Altschul, Phys. Rev. D **70**, 056005 (2004); G. Shore, hep-th/0409125.
52. B. Heckel, in *CPT and Lorentz Symmetry III*, [12]; L.-S. Hou, W.-T. Ni, and Y.-C.M. Li, Phys. Rev. Lett. **90**, 201101 (2003); R. Bluhm and V.A. Kostelecký, Phys. Rev. Lett. **84**, 1381 (2000).
53. H. Müller, S. Herrmann, A. Saenz, A. Peters, and C. Lämmerzahl, Phys. Rev. D **68**, 116006 (2003); Phys. Rev. D **70**, 076004 (2004).
54. V.W. Hughes, H.G. Robinson, and V. Beltran-Lopez, Phys. Rev. Lett. **4** (1960) 342; R.W.P. Drever, Philos. Mag. **6** (1961) 683; J.D. Prestage et al., Phys. Rev. Lett. **54** (1985) 2387; S.K. Lamoreaux et al., Phys. Rev. A **39** (1989) 1082; T.E. Chupp et al., Phys. Rev. Lett. **63** (1989) 1541; C.J. Berglund et al., Phys. Rev. Lett. **75** (1995) 1879; D. Bear et al., Phys. Rev. Lett. **85**, 5038 (2000);
55. D.F. Phillips et al., Phys. Rev. D **63**, 111101 (2001); M.A. Humphrey et al., Phys. Rev. A **68**, 063807 (2003); Phys. Rev. A **62**, 063405 (2000).
56. F. Canè et al., Phys. Rev. Lett. **93**, 230801 (2004).
57. V.A. Kostelecký and C.D. Lane, Phys. Rev. D **60**, 116010 (1999); J. Math. Phys. **40**, 6245 (1999).
58. R. Bluhm et al., Phys. Rev. Lett. **88**, 090801 (2002); Phys. Rev. D **68**, 125008 (2003).
59. G. Gabrielse et al., Phys. Rev. Lett. **82** (1999) 3198.
60. KTeV Collaboration, H. Nguyen, in *CPT and Lorentz Symmetry II*, [12]; OPAL Collaboration, R. Ackerstaff et al., Z. Phys. C **76**, 401 (1997); DELPHI Collaboration, M. Feindt et al., preprint DELPHI 97-98 CONF 80 (1997); BELLE Collaboration, K. Abe et al., Phys. Rev. Lett. **86**, 3228 (2001); BaBar Collaboration, B. Aubert et al., Phys. Rev. Lett. **92**, 142002 (2004); FOCUS Collaboration, J.M. Link et al., Phys. Lett. B **556**, 7 (2003).
61. V.A. Kostelecký, Phys. Rev. Lett. **80**, 1818 (1998); Phys. Rev. D **61**, 016002 (2000); Phys. Rev. D **64**, 076001 (2001).
62. V.W. Hughes et al., Phys. Rev. Lett. **87**, 111804 (2001).
63. H.N. Brown et al., Phys. Rev. Lett. **86**, 2227 (2001).
64. R. Bluhm, V.A. Kostelecký and C.D. Lane, Phys. Rev. Lett. **84**, 1098 (2000).
65. Recent experimental studies of Lorentz and CPT violation with neutrinos are summarized in papers by M.D. Messier (SK); T. Katori and R. Tayloe (LSND); and B.J. Rebel and S.F. Mufson (MINOS); all in *CPT and Lorentz Symmetry III*, [12].
66. V.A. Kostelecký and M. Mewes, Phys. Rev. D **69**, 016005 (2004); Phys. Rev. D **70**, 031902(R) (2004); Phys. Rev. D **70**, 076002 (2004).
67. V. Barger, S. Pakvasa, T. Weiler, and K. Whisnant, Phys. Rev. Lett. **85**, 5055 (2000); J.N. Bahcall, V. Barger, and D. Marfatia, Phys. Lett. B **534**, 114 (2002); I. Mocioiu and M. Pospelov, Phys. Lett. B **534**, 114 (2002); A. de Gouvêa, Phys. Rev. D **66**, 076005 (2002); G. Lambiase, Phys. Lett. B **560**, 1 (2003); S. Choubey and S.F. King, Phys. Lett. B **586**, 353 (2004); A. Datta et al., Phys. Lett. B **597**, 356 (2004).

68. D.L. Anderson, M. Sher, and I. Turan, Phys. Rev. D **70**, 016001 (2004); E.O. Iltan, Mod. Phys. Lett. A **19**, 327 (2004).
69. L. Lavoura, Ann. Phys. **207**, 428 (1991).
70. KTeV Collaboration, Y.B. Hsiung et al., Nucl. Phys. Proc. Suppl. **86**, 312 (2000); B. Winstein, in *CPT and Lorentz Symmetry II*, [12].
71. Additional terms that are forbidden by the requirements of gauge invariance and renormalizability in the minimal SME can arise effectively in a QED extension due to strong binding in a nucleus. These coefficients are labeled as e_μ, f_μ, and $g_{\lambda\mu\nu}$. The are included in the general investigations described in [57]. However, for simplicity, they are ignored here.
72. B. Heckel, in *CPT and Lorentz Symmetry III*, [12].

Anything Beyond Special Relativity?

G. Amelino-Camelia

Dipartimento di Fisica, Universitá di Roma "La Sapienza" and Sez. Roma1 INFN, P.le Moro 2, I-00185 Roma, Italy
amelino@roma1.infn.it

At a time when we are all fully aware of the remarkable longevity of Special Relativity, I review some recent advances in "beyond Special Relativity" research, attempting to make the case that this field is reaching a fair level of maturity. I briefly discuss both some recent attempts to set up systematic general searches of departures from Special Relativity, and some more focused programmes which find their motivation in certain key aspects of the quantum-gravity problem and in certain open issues for cosmology. For one of the hypothesis being considered in the quantum-gravity literature, the one of Planck-scale effects that, while inducing departures from some Special-relativistic laws, do not give rise to a loss of equivalence among inertial observers, I give a rather detailed self-contained introduction. I also stress the fact that beyond-Special-Relativity research is now being developed with a methodology that in some ways resembles the one adopted in the "beyond the Standard Model" particle-physics research programme. I argue that the rich tradition of beyond-Standard-Model research should be followed even more closely, but I also stress that some differences will inevitably remain, reflecting the differences between setting up test theories for a "universal"/"frame" theory (Special Relativity) and for a nonuniversal theory (the Standard Model).

1 Introduction and Summary

Beyond Special Relativity

One might say that Special Relativity is not really a "fundamental" theory: it only emerges in a particular limit (a specific solution of the equations) of General Relativity. But Special Relativity, now 100 years old, still is legitimately viewed as fundamental in a wide class of physical contexts. Even in analyses involving gravitational phenomena one often gets away describing the spacetime metric $g_{\mu\nu}$ in terms of a Minkowski background metric, $\eta_{\mu\nu}$, and a "gravity field" Lorentz tensor $h_{\mu\nu}$, related to g and η by the relation $h_{\mu\nu} = g_{\mu\nu} - \eta_{\mu\nu}$. And special relativity reigns supreme in the vast class of phenomena studied

G. Amelino-Camelia: *Anything Beyond Special Relativity?*, Lect. Notes Phys. **702**, 227–278 (2006)
DOI 10.1007/3-540-34523-X_9 © Springer-Verlag Berlin Heidelberg 2006

in particle physics, where one can safely assume that the processes unfold in a Minkowski background spacetime, with metric $\eta_{\mu\nu}$, and that "gravitational interactions" among particles are negligible.

Of course as the success of Special Relativity becomes more and more remarkable scientists become more and more determined to place it under further scrutiny. Indeed over these past few years there has been a renewed effort in setting up formalisms suitable for use in exploring systematically possible departures from special-relativistic laws. And, besides these necessary "look-everywhere tests" of Special Relativity, in these past few years there has also been increased interest in the possibility that departures from certain relevant special-relativistic laws might lead to the solution of some outstanding open issues for theoretical physics.

Quantum Gravity (and Cosmology)

In particular, several authors have argued that the transition from (special-) relativistic quantum field theory to (the still unknown) "quantum gravity" might force special relativity to relinquish even its present privileged status within particle physics, as soon as we acquire sensitivity to Planck-scale corrections to particle-physics processes.

The description of Planck-scale corrections to particle-physics processes will be a key aspect of the Minkowski limit of quantum gravity. In our current conceptual framework Special Relativity emerges in the Minkowski limit, where one deals with situations that allow the adoption of a Minkowski metric throughout, and one might wonder whether the Minkowski limit of quantum gravity could still be governed by Special Relativity. The issue will be of particular interest if quantum gravity admits a limit in which one can assume throughout a (expectation value of the) metric of Minkowski type, but some Planck-scale features of the fundamental description of spacetime (such as spacetime discreteness and/or spacetime noncommutativity) are still not completely negligible. I will denominate "nontrivial Minkowski limit" this type of Minkowski limit in which essentially the role of the Planck scale in the description of gravitational interactions (expressing the gravitational constant G in terms of the Planck scale) can be ignored, but the possible role of the Planck scale in spacetime structure/kinematics is still significant. For various approaches to the quantum-gravity problem evidence as emerged in support of this possibility of a nontrivial Minkowski limit. While there is no fully-developed proposed solution of the quantum-gravity problem based on a fundamentally noncommutative spacetime picture, it has been observed that the hypothesis that in general the correct fundamental description of spacetime should involve noncommutativity can imply that in particular the Minkowski limit is described in terms of noncommuting spacetime coordinates, and this is found [1–8] to naturally lead to a nontrivial Minkowski limit with departures from classical Poincaré symmetry. In the literature on the loop-quantum-gravity approach one finds a large number (although all of preliminary nature) of arguments [9–13] supporting the possibility of a nontrivial Minkowski limit, primarily characterized by a Planck-scale-modified

energy-momentum (dispersion) relation. For the string-theory approach, while there are no studies arguing that the availability of a nontrivial Minkowski limit is necessary, there is a large literature (see, e.g., [5,6] and references therein) on a nontrivial Minkowski limit with broken Lorentz symmetry.

The fate of Poincaré symmetry in such nontrivial Minkowski limits is of course a key issue both phenomenologically and from a conceptual perspective. In the large number of studies produced between 1997 and 2000 on the possibility of a nontrivial Minkowski limit for quantum gravity it was always assumed that Poincaré (and in particular Lorentz) symmetry would be broken: the Galilei Relativity Principle would not hold with Planck-scale accuracy. On the basis of an analogy with the century-old process which led from Galilei/Newton Relativity, through the analysis of Maxwell's electrodynamics,[1] to Einstein's Special Relativity, I argued in [14] that the Minkowski limit of quantum gravity might be characterized by a "doubly special relativity" (DSR), a relativistic theory with two, rather than one, nontrivial relativistic invariants (the Planck scale in addition to the speed-of-light scale), but still fully compatible with the Galilei Relativity Principle. Since the idea of a Planck-scale-broken Poincaré symmetry has a long tradition in the quantum-gravity literature, and should be familiar to most readers, whereas the hypothesis of a "DSR Minkowski limit" has only been considered over these past few years, I will provide in these lectures a self-contained introduction to DSR, but I will only comment briefly on Planck-scale broken-Poincaré-symmetry scenarios.

I should stress that besides the study of the quantum-gravity problem there are other research areas in which the possibility of some departures from special-relativistic laws is being considered as a possible solution to some outstanding open issues. In particular, it has been observed [15–17] that the hypothesis of a "time-varying speed-of-light constant"[2] could address some of the same issues that in cosmology are usually described in terms of inflation. Just like with inflation one places far-away regions of the Universe in causal contact by introducing an "accelerated expansion era", a law for the time-variation of the speed-of-light constant in which the speed of light has a larger value at earlier times would also place in causal contact some far-away regions of the Universe. Since inflation-based cosmology still lacks a genuine "smoking gun" data verification, and there are some researchers finding some aspects of inflation not fully satisfactory at the conceptual level, it is not surprising that time-varying speed-of-light cosmology is attracting some interest. Moreover, there is some tentative evidence [18,19] of time-varying "constants" (such as α) which may in turn be used [17] to motivate research on a time-varying speed-of-light constant.

[1] The role of the speed scale "c" in Maxwell's electrodynamics was first viewed as a manifestation of a ether violating the Relativity Principle, but was ultimately understood as a manifestation of a needed transition in the formulation of the relativistic theory.

[2] I follow other authors in speaking of "time-varying constants"...of course if their values did indeed change in time they would not be constants... but they used to be considered as constants...

Concerning the possible encouragement from preliminary data, the situation for time-varying speed-of-light cosmology is actually rather analogous to the one of Planck-scale departures from special relativity within quantum gravity: with presently-available data one could argue that the predicted "GZK cutoff" for cosmic rays is not being seen [18,19], and a possible explanation could come [20–23] from Planck-scale departures from special relativity.

Planck-scale departures from special relativity have actually been advocated in a recent variant of time-varying speed-of-light cosmology. One of the effects most discussed in the quantum-gravity literature on Planck-scale departures from special relativity is a Planck-scale modification of the energy-momentum (dispersion) relation, with an associated energy-dependence of the speed of massless particles. In some cases this energy dependence is such that the speed can grow very large at high energies, and since in the early Universe the typical energies of particles were very high, this scenario would also provide an "effective time dependence" of the speed of light: there would actually be no genuine time dependence, but the Planck-scale-induced energy dependence of the speed of photons would mimic a time dependence, since at different times in the evolution of the Universe particles are expected to have different typical energies.

Just Like "Beyond the Standard Model"

Very few authors, especially in science magazines, appear to notice the similarities between beyond-Special-Relativity research and the "beyond the Standard Model" particle-physics research programme. Somehow the success of Special Relativity is often foolishly perceived as an indication that "this theory is right",[3] while the success of the Standard Model particle physics is correctly viewed as an indication that this theory is somehow very accurate in the regimes presently accessible to us, but should still eventually break down. Seminars contemplating possible departures from Special Relativity often start with one form or another of apology, while no apology is offered at seminars considering grandunification, supersymmetric gauge theories, technicolor models...

There are of course some reasons for this. First of all it must be acknowledged that, especially in the "old days", there have been too many papers that were claiming to look "beyond Special Relativity", but actually only reflected poor understanding of Special Relativity itself. And there have been too many papers in which the motivation proposed for the analysis did not go much further than Star Trek. Perhaps more importantly, departures from Special Relativity have been already "discovered" many times by authors who however did not realize fully, e.g., the difference between phase velocity and signal velocity. It is then not surprising that for a portion of the physics community (and an even more significant portion of the community of outsiders) the motivations for research beyond the Standard Model of particle physics appear clear and robust, while a cloud of suspicion surrounds research beyond Special Relativity.

[3] I write this in quotes since this sentence (in spite of being frequently used) has no meaning.

Still, as beyond-Special-Relativity research is becoming more mature, it is legitimate to expect that the perception will also change. After all both beyond-Special-Relativity and beyond-Standard-Model research are primarily driven by hints that point us toward additional structures in the laws of Nature at short-distance/high-energy scales. It is natural to guide the experimental tests in the direction of some open conceptual issues, such as the "hierarchy problem" on one side and the short-distance structure of spacetime on the other side. And, since indeed both the Standard Model and Special Relativity are in excellent agreement with all presently-available data, the primary motivation for beyond-Special-Relativity and beyond-Standard-Model research is just to provide guidance for further testing of these remarkably successful theories. So far both have succeeded: guidance has been provided to those performing experimental tests (but all the results of the tests have been negative on both sides).

While some similarities between these two research programmes are a natural consequence of their similar objectives, there are even some other, apparently fortuitous, similarities. For example, it is amusing to notice that in the mentioned "doubly special relativity" part of beyond-Special-Relativity research one of the most studied candidate formalisms is based on a Hopf-algebra description of spacetime symmetries, and Hopf algebras may be viewed as a sort of "loophole" of the Coleman-Mandula theorem, just like the idea of supersymmetry, the most popular beyond-Standard-Model research programme. The Coleman-Mandula theorem, focusing on symmetry algebras that are fully characterized by commutators of the generators, basically leads to the conclusion that in particle physics there cannot be alternatives to (or extensions/generalizations of) the Poincaré symmetry algebra. In a Hopf-algebra formulation of the description of spacetime symmetries one contemplates the possibility that the symmetry generators are also to satisfy nontrivial co-commutator relations (see later), while supersymmetric particle-physics models are based on symmetry generators which also satisfy nontrivial anti-commutator relations.[4]

There are of course also some important differences between these two research programmes, due to the fact Special Relativity is a "universal" (or "frame") theory [25], a theory applicable to any kind of physical phenomenon, whereas the Standard Model of particle physics is a "nonuniversal" theory, which is intended only for the description of one aspect (some of the interactions among particles) of the particle-physics arena. Indeed the Standard Model is subject to Special Relativity, and all successful tests of the Standard Model may also be viewed as tests of Special Relativity (while the reverse of course is not true). This also affects the development of test theories: in considering departures from Special Relativity one often ends up facing the challenge of developing completely new formalisms that introduce significant changes in the "rules of the game", instead particle-physics theories "beyond the Standard Model" typically still make use of the familiar machinery based on relativistic quantum field theory, gauge invariance and all that. Perhaps it is for this reason that in looking

[4] I have heard this point about the Coleman-Mandula theorem, supersymmetry algebras and Hopf algebras most elegantly stressed in seminars by J. Wess (e.g., [24]).

beyond Special Relativity one often cannot avoid feeling disoriented, as if one was looking in the dark, while at least some ways to introduce departures from the Standard Model of particle physics allow us to work with the reassuring feeling of a systematic approach within a known scheme.

Plan of These Notes

In the next section I describe briefly some aspects of beyond-Special-Relativity research, focusing on research programmes that attempt to set up a systematic investigation of some classes of possible departures from Poincaré symmetry, on research programmes that are inspired by the study of the quantum-gravity problem, and on research programmes that are inspired by some open issues in cosmology.

Then, in Sect. 3, I provide a more detailed description of some quantum-gravity scenarios which would indeed lead to departures from Poincaré symmetry. And, in Sect. 4, I give a brief, but self-contained, introduction to "doubly-special relativity".

In Sect. 5, I provide some additional comments on the similarities between beyond-Special-Relativity and beyond-Standard-Model research, while in Sect. 6 I offer some closing remarks.

2 Some Key Aspects of Beyond-Special-Relativity Research

The primary motivation of research beyond Special Relativity is of course the one of providing targets for the experimentalists that are testing Special Relativity. And one is naturally tempted to set up formalism encoding "all possible departures" from Special Relativity, so that the experimental searches could be as general as possible and the formalism would provide to experimentalists a sort of "language" to use in comparing experimental results. While it is indeed important to be guided by this objective, in practice the concept of a formalism encoding "all possible departures" from Special Relativity is not even well defined: one may allow a large number of free parameters, but in setting up the formalism one inevitably ends up making assumptions about the form of the laws that one is contemplating. It is therefore important to also develop focused research programmes that look at some open theoretical-physics issues and explore the possible implications for Special Relativity, since this type of exercise will sometimes expose some hidden assumptions in other approaches. For example, the mentioned idea of a "doubly special relativity" made us realize that previous test theories for Special Relativity were only testing scenarios in which not only the second postulate but also the first "Relativity Principle" postulate would be violated. We now know that this is not the only option: there might be some departures from the second postulate, while still preserving the validity of the first postulate (i.e. preserving the equivalence of inertial observers). Another

relevant example, coming from outside research beyond Special Relativity (but closely-related research testing another symmetry), is the one of test theories for CPT symmetry, which usually assume departures from CPT symmetry that are independent of the momenta of the particles entering the process. Some quantum pictures of spacetime ended up motivating the presence of some momentum dependence [26] and whereas within the momentum-independent scheme the best presently-available limits are established using neutral-kaon data (see, e.g., [27]), the analysis of some momentum-dependent departures from CPT symmetry suggests [26] that neutral-B-meson experiments should be most sensitive.

In summary, while it would be nice to claim that all possible departures from Special Relativity are being systematically investigated, in practice we combine some simple general schemes of wide applicability, which provide some overall guidance to tests of Special Relativity and make "the minimum possible" number of assumptions, with some programmes looking carefully at open issues in theoretical physics, looking for hints of specific departures from Special Relativity, possibly also of types not previously included in the general schemes guiding the first-level "systematic" explorations.

As open issues in theoretical physics which may provide hints toward physics beyond Special Relativity I consider here as "good examples" the one of research on the quantum-gravity problem and research on alternatives to the inflation mechanism in cosmology. But before going to these topics, let me start by describing briefly some attempts of "general formalisms for systematic searches of departures from Special Relativity".

Some General Formalisms for Systematic Searches

A popular attempt to provide a framework for systematic searches of departures from Special Relativity is the so-called "Standard Model Extension" (see, e.g., [28–32]), which allows for a very general parametrization of new effects within the context of a Lagrangian-based quantum-field-theory setup. Besides the strict implementation of the field-theory machinery this approach assumes various other building principles, including classical conservation of energy-momentum, Hermiticity, microcausality, positivity of energy, gauge invariance, and power-counting renormalizability.[5] Moreover it is also assumed that the departures from Special Relativity would be of a type that one could possibly describe in a spontaneous-symmetry-breaking setup: the Lagrangian is a Lorentz-transformation scalar and the departures from Special Relativity are only obtained by assuming that some of the vectors and tensors that appear in the Lagrangian are not describing dynamical fields but are rather fixed/nondynamical

[5] I am calling "Standard Model Extension" the model described with this name in [28, 29] Readers should however notice that in the most recent literature (see, e.g., [33]) this same model is sometimes referred to as "minimal Standard Model Extension", and instead with "Standard Model Extension" one denotes a generalization which allows for powercounting-nonrenormalizable terms, while insisting on the other premises of the original Standard Model Extension.

tensors[6] (which may well be the vacuum expectation value of some dynamical field). Essentially, the parameters of the Standard Model Extension are just additional interactions with constant fields.[7]

The Standard Model Extension is an excellent example of what one would naturally do in trying to provide a general framework for systematic searches of departures from Special Relativity: it introduces a very general parametrization, able to describe a large variety of new effects, and relies on a well-defined and rather reasonable set of assumptions about the "rules of the game". Essentially the Standard Model Extension assumes that the departures from Special Relativity could be described by adding new parameters, but no conceptually new structures, to the presently-adopted formalisms. This is reassuring from a theoretical perspective, since the logical consistency of the approach is relatively safe (it relies on the logical consistency of the original theories), and is advantageous for experimentalists, since it allows them to interpret their data within the context of a familiar formalism. Still, it may well be that the correct description of departures from Special Relativity would instead require the introduction of new structures, a change in the rules of the game. For example, most of the quantum-gravity intuition (see later) for departures from Special Relativity invites one to consider effects that are not described by power-counting-renormalizable terms. And indeed some authors, even when adopting an approach that is primarily based on quantum field theory, have chosen to look beyond the Standard-Model-Extension setup, considering Planck-scale-suppressed effects which are in fact not described by power-counting-renormalizable terms (see, e.g., [36, 37]).

Of course, it is also possible to renounce to the assumption of a Lagrangian generating the dynamical equations, and in fact there is a rich phenomenology being developed introducing the generalizations directly at the level of the dynamical equations [38,39]. This is of course more general than the Lagrangian approach: for example, the generalized Maxwell equation discussed in [39] predicts effects that go beyond the Standard Model Extension. And charge conservation, which automatically comes out from the Lagrangian approach, can be violated in models generalizing the field equations [39]. However, if one "by hand" generalizes the dynamical equations it is necessary to proceed cautiously [39, 40] in order to ensure the logical consistency of the approach.

The comparison of the Standard-Model-Extension approach and of the approach based on generalizations introduced directly at the level of the dynamical equations illustrates how different "philosophies" lead to different strategies for setting up a "completely general" systematic investigation of possible departures from Special Relativity. By removing the assumption of the availability of

[6] Of course, I am considering here fixed tensors, which are tensors that take different (matrix) value in different reference frames. If the vacuum of a field theory is characterized by a fixed (nondynamical) tensor then different observers are not equivalent, since they can be distinguished from one another using indeed the fact that the tensor takes different values in different inertial frames.

[7] This kind of generalizations in the photonic sector had been introduced and discussed earlier by Ni [34] and Haugan and Kauffmann [35].

a Lagrangian the second approach is "more general". Still no "general approach" can be absolutely general: in principle one could always consider removing an extra layer of assumptions.

Some quantum-gravity arguments actually do provide motivation to remove even more assumptions. A significant portion of the quantum-gravity community is in general, justifiably, skeptical about the results obtained using low-energy effective field theory in analyses relevant for the quantum-gravity problem. After all the first natural prediction of low-energy effective field theory in the gravitational realm is a value of the energy density which is some 120 orders of magnitude greater than allowed by observations.[8] This observation may be combined with various arguments that suggest the possibility of Planck-scale-induced departures from quantum mechanics ("quantum-gravity-induced decoherence"), which also should render us cautious about applying the standard description of dynamics, through quantum field theory and the dynamical equations of motion, in the quantum-gravity realm. Moreover, it has emerged (see later) that in some quantum-gravity scenarios, if one goes ahead anyway applying the effective-field-theory machinery, one stumbles upon a "IR/UV mixing" (see, e.g., [5,6]) which basically implies that low-energy effective field theory, when applied in those quantum-gravity contexts, is void of any predictive power [41]. Basically this theories do not enjoy Wilson decoupling: whereas our ignorance of the UV sector of the laws of Nature does not affect the low-energy predictivity of theories that enjoy Wilson decoupling, in presence of "IR/UV mixing" one can obtain a reliable description of low-energy physics only when the laws of Nature are exactly known all the way up to the infinite-energy regime.

While of course we cannot exclude that somehow low-energy-effective-field-theory techniques be applicable to the quantum-gravity realm, in light of the information presently available to us it is rather naive to assume that our present standard description of dynamics should necessarily work in the description of Planck-scale-induced effects. If the arguments that encourage the use of new descriptions of dynamics at the Planck scale are correct, then a sort of "order of limits problem" clearly arises. We know that in some limit (a limit that characterizes our most familiar observations) the field-theoretic description and Lorentz invariance will hold. So we would need to establish whether experiments that are sensitive to Planck-scale departures from Lorentz symmetry could also be sensitive to Planck-scale departures from the field-theoretic description of dynamics. As an example, one may consider the possibility (not unlikely in a context which is questioning the fate of Lorentz symmetry) that quantum gravity would admit a field-theory-type description only in reference frames in which the process of interest is essentially occurring in its center of mass (no "Planck-large

[8] And the outlook of low-energy effective field theory in the gravitational realm does not improve much through the observation that exact supersymmetry could protect from the emergence of any energy density. In fact, Nature clearly does not have supersymmetry at least up to the TeV scale, and this would still lead to a natural prediction of the cosmological constant which is some 60 orders of magnitude too high.

boost" [42] with respect to center-of-mass frame). The field theoretic description could emerge in a sort of "low-boost limit", rather than the expected low-energy limit.[9]

As a result of these concerns for the description of dynamics, several authors have adopted an approach to the search of Planck-scale-induced departures from Special Relativity which focuses on pure kinematics. This of course restricts the number of experimental contexts from which to obtain experimental limits, since it is not easy to find situations [43] in which a pure-kinematics analysis is possible (often even some aspects which at first appear to be relevant only for kinematics are actually affected indirectly by the description of dynamics), but it eliminates some potentially unreliable assumptions about the description of dynamics.

Most of the recent work[10] within this pure-kinematics approach may be primarily characterized by the following parametrization of the energy-momentum (dispersion) relation

$$m_a^2 \simeq E^2 - \mathbf{p}^2 + \gamma_a \mathbf{p}^2 + \eta_a \mathbf{p}^2 \frac{E}{E_p} + \zeta_a \mathbf{p}^2 \frac{E^2}{E_p^2} + \dots \ , \tag{1}$$

where E_p denotes as usual the Planck energy scale ($\sim 10^{28}\,\mathrm{eV}$) and the index "$a$" leaves room for a possible dependence of the effects on the type of particle which is being considered.

While in the first studies [9, 36] that proposed a phenomenology based on (1) the key effect under consideration concerned the "signal velocity", obtained from the dispersion relation according to[11] $v = dE/dp$, more recently (starting

[9] The regime of low boosts with respect the center-of-mass frame is often indistinguishable with respect to the low-energy limit. For example, from a Planck-scale perspective, our laboratory experiments (even the ones conducted at, e.g. CERN, DESY, SLAC...) are both low-boost (with respect to the center of mass frame) and low-energy. However, the "UHE cosmic-ray paradox", for which a quantum-gravity origin has been conjectured [20, 23], occurs in a situation where all the energies of the particles are still tiny with respect to the Planck energy scale, but the boost with respect to the center-of-mass frame (as measured by the ratio E/m_{proton} between the proton energy and the proton mass) could be considered to be "large" from a Planck-scale perspective ($E/m_{proton} \gg E_p/E$, with E_p denoting the Planck energy scale).

[10] Although recently most of the activity on the "pure-kinematics front" takes as motivation the quantum-gravity problem and is set up in such a way to reflect that intuition, the general idea of pure-kinematics Special-Relativity test theories has a tradition that extends over several decades and was not originally connected with quantum-gravity research. In particular, the RMS (Robertson–Mansouri–Sexl) test theory [44,45], which introduces anomalous effects for light propagation, has received significant consideration by experimentalists.

[11] Even the apparently safe assumption of "$v = dE/dp$" has been, and understandably, challenged by some quantum-gravity studies (see, e.g., [46–49]). As stressed in [43,50] one possible concern here is whether quantum gravity leads to a modified Heisenberg uncertainty principle, $[x, p] = 1 + F(p)$. Assuming a Hamiltonian description is still available, $v = dx/dt \sim [x, H(p)]$, the relation $v = dE/dp$ essentially follows from

with the studies reported in [20–23, 51]) the phenomenology has considered also the possibility of Planck-scale modifications of the kinematical thresholds for particle reactions, and the analyses also relied on some general parametrization of possible Planck-scale departures from the law of energy-momentum conservation.

This concludes my list of four possible "completely general" (each not really completely general) approaches to the search of departures from Special Relativity: (a) the Standard Model Extension, (b) the Standard Model Extension enriched by power-counting-nonrenormalizable terms, whose phenomenology was first advocated in [9, 36], (c) the approach based on direct modification of dynamical equations (without necessarily demanding that the modified equation be derivable from a Lagrangian), and (d) the pure-kinematics approach. I shall not dwell on whether or not this is a complete list: my true objective was just to expose the fact that there are many "completely general" approaches (characterized by different assumptions), and therefore none of them is really completely general.

Intuition from Quantum Gravity

Already in my discussion of "general formalisms for systematic searches" I mentioned some elements of intuition that have originated in the quantum-gravity literature: the possibility of effects whose magnitude is governed by ratios of the energy of the particles involved in the processes and the Planck energy scale (which in field-theory language invites one to consider terms which are not power-counting renormalizable) and the possibility that departures from Special Relativity might accompanied by departures from our presently-adopted laws of description of dynamics. But this is only a very small sample of the elements of intuition being provided by quantum-gravity research. The variety of scenarios for the fate of Special Relativity at the Planck scale is so large that I am setting aside for it the next section.

Let me just anticipate here the fact that some quantum-gravity scenarios (or quantum-spacetime pictures) also invite to consider the possibility of energy-dependent birefringence: photons may propagate in a way that depends on their direction of polarization as seen by a certain preferred frame and the magnitude of this effect may depend on the energy (wavelength) of the photon.

Moreover, as already mentioned in the Introduction, while all non-quantum-gravity-motivated tests of Special Relativity actually tested the possibility of "broken Poincaré symmetry" (departures from Special Relativity that are such that the equivalence of inertial observers is lost), the "doubly-special relativity" scenario, which emerged recently [14, 52–56] in the quantum-gravity literature, provides motivation for tests of the possibility of "Planck-scale deformed

$[x, p] = 1$. But if $[x, p] \neq 1$ then $v = dx/dt \sim [x, H(p)]$ would not lead to $v = dE/dp$. There is much discussion in the quantum-gravity community of the possibility of modifications of the Heisenberg uncertainty principle at the Planck scale, and this would invite us once again to remove an additional layer (the "$v = dE/dp$" layer) of assumptions inherited from the standard formalism, but I shall not dwell on this possibility in these notes.

Poincaré symmetry". Section 4 is devoted to a self-contained introduction of this scenario.

Intuition from Cosmology

There is research beyond Special Relativity also in cosmology. There the motivation usually comes from two sources, a theory intuition and some preliminary data. On the theory side, some authors, in spite of the good level of success of inflation models, are still concerned about some conceptually unappealing aspects of inflation, such as the high level of fine tuning that inflation models require. And inflation is primarily considered as a way to introduce an early-Universe time interval in which effectively the "causality light cone" grows very large, placing in causal contact regions of the Universe that would otherwise be disconnected. Some authors have considered the possibility that the light cone might have really been larger at early times, by introducing [15–17] a "time-varying speed-of-light constant". If the time dependence is adjusted appropriately, this scenario can of course replace inflation in creating a framework for placing in causal contact regions of the Universe that would instead be disconnected if the speed of light was time independent.

On the experiment side, the general idea of time-varying "constants", and in particular a time-varying speed-of-light constant, receives some (however preliminary) encouragement from observations [18, 19] which could indeed suggest a time variation of time-varying "constants" (such as α).

These arguments from the cosmology literature may provide motivation for a specific type of tests of Special Relativity, in which somehow the basic relativistic structures are largely preserved, but the constant c is replaced by a time-dependent $c(t)$.

As mentioned in the first section, there is also a certain level of overlap between the quantum-gravity research beyond Special Relativity and cosmology research beyond Special Relativity, as illustrated by the very recent scenarios [17,57] in which Planck-scale departures from Special Relativity (possibly introducing a wavelength dependence of the speed of photons) are used to obtain a new variant of time-varying-speed-of-light cosmology (since in the early Universe the typical wavelength of particle was much higher than at present times). And other areas of overlap between these two motivations for research beyond Special Relativity may soon emerge: for example, in Loop Quantum Gravity there is much discussion of possible departures from some special-relativistic laws [9–13] and there has also been a strong interest in the implications of Loop Quantum Gravity for cosmology [58–60]. If indeed Loop Quantum Gravity hosts some departures from Special Relativity it is likely that they would affect significantly the analyses on the cosmology side.

Aside on Analog Models and Condensed-Matter Systems

I choose not to comment on any other branch of research beyond Special Relativity, which could provide (like quantum gravity and cosmology) inspiration

for focused searches of some specific departures from Special Relativity. There are some other proposals which are guided by defendable motivation, but I find more useful for the scopes of these notes to mention, however briefly, a research line in which some condensed-matter models are used to get intuition on how Special Relativity might be violated at high energies. Most of the key tests of Special Relativity involve particle physics, but particle physicists typically only have familiarity with certain specific mechanisms (for example, the spontaneous-symmetry-breaking mechanism) to introduce departures from a given symmetry. Condensed-matter theorists have a complementary expertise in dealing with symmetries. They are used to describe the degrees of freedom that are measured in the laboratory as collective excitations within a theoretical framework whose primary description is given in terms of much different, and often practically inaccessible, fundamental degrees of freedom. And in some regimes (e.g. close to a critical point) some symmetries arise for the collective-excitations theory, which do not carry the significance of fundamental symmetries, and are in fact lost as soon as the theory is probed somewhat away from the relevant regime. Notably, some familiar systems are known to exhibit a form of special-relativistic invariance in certain limits, even though, at a more fundamental level, they are described in terms of a nonrelativistic theory.

Besides the intuition emerging form the study of open problems in quantum-gravity research and cosmology, it is important that the searches of possible departures from Special Relativity take into account the intuition coming from well-understood (but little-known within the particle-physics community) condensed-matter systems which may be viewed as toy models for spacetime physics. Among these possible toy models an important class is the one of "gravity analog models" (see, e.g., [61–63]). For example, there is close analogy between the propagation of sound waves in a moving fluid and the propagation of light waves in a curved spacetime. And there is a possible connection between the "analog-model intuition" and the "quantum-gravity intuition", as illustrated by the observation [61] that supersonic fluid flow can generate a "dumb hole", the acoustic analogue of a black hole, and one can study "phononic Hawking radiation" from the acoustic horizon.

3 More on the Quantum-Gravity Intuition

After a rather general perspective adopted in the first pages of these notes, I am now going to gradually focus the analysis by first discussing how specifically the quantum-gravity problem motivates the study of departures from Special Relativity, and then, in the next section, I will consider one particular scenario, the "doubly-special relativity" scenario, for Planck-scale departures from Special Relativity.

In this section I first discuss how some alternative perspectives on the quantum-gravity problem provide different types of intuition toward the fate of Special Relativity at the Planck scale. Then I comment on departures from

Special Relativity which one encounters in the study of the most popular approaches to the quantum-gravity problem: string theory, loop quantum gravity, and noncommutative geometry.

3.1 The Three Perspectives on the Quantum-Gravity Problem and Their Implications for the Fate of Special Relativity at the Planck Scale

It is probably fair to state that each quantum-gravity research line can be connected with one of three perspectives on the problem: the particle-physics perspective, the General-Relativity perspective and the condensed-matter perspective.

From a particle-physics perspective it is natural to attempt to reproduce as much as possible the successes of the Standard Model of particle physics. One is tempted to see gravity simply as one more gauge interaction. From this particle-physics perspective a natural solution of the quantum-gravity problem should have its core features described in terms of graviton-like exchange in a background classical spacetime. Indeed this structure is found in String Theory, the most developed among the quantum-gravity approaches that originate from a particle-physics perspective.

The General-Relativity perspective naturally leads us to reject the use of a background spacetime [64, 65]. According to General Relativity the evolution of particles and the structure of spacetime are selfconsistently connected: rather than specify a spacetime arena (a spacetime background) beforehand, the dynamical equations determine at once both the spacetime structure and the evolution of particles. Although less publicized, there is also growing awareness of the fact that, in addition to the concept of background independence, the development of General Relativity relied heavily on the careful consideration of the in-principle limitations that measurement procedures can encounter.[12] In light of the various arguments suggesting that, whenever both quantum mechanics and General Relativity are taken into account, there should be an in-principle Planck-scale limitation to the localization of a spacetime point (an event), the general-relativity perspective invites one to renounce to any direct reference to a classical spacetime [66–70]. Indeed this requirement that spacetime be described as fundamentally nonclassical ("fundamentally quantum"), the requirement that the in-principle measurability limitations be reflected by the adoption of a corresponding measurability-limited description of spacetime, is another element of intuition which is guiding quantum-gravity research from the general-relativity perspective. This naturally leads one to consider discretized spacetimes, as in the Loop Quantum Gravity approach, or noncommutative spacetimes.

The third possibility is a condensed-matter perspective on the quantum-gravity problem (see, e.g., [71, 72]), in which spacetime itself is seen as a sort

[12] Think for example of the limitations that the speed-of-light limit imposes on certain setups for clock synchronization and of the contexts in which it is impossible to distinguish between a constant acceleration and the presence of a gravitational field.

of emerging critical-point entity. As stressed already in the previus section, condensed-matter theorists are used to describe the degrees of freedom that are measured in the laboratory as collective excitations within a theoretical framework whose primary description is given in terms of much different, and often practically inaccessible, fundamental degrees of freedom. Close to a critical point some symmetries arise for the collective-excitations theory, which do not carry the significance of fundamental symmetries, and are in fact lost as soon as the theory is probed somewhat away from the critical point. Notably, some familiar systems are known to exhibit special-relativistic invariance in certain limits, even though, at a more fundamental level, they are described in terms of a nonrelativistic theory.

Clearly for the condensed-matter perspective on the quantum-gravity problem it is natural to see the familiar classical continuous Lorentz symmetry only as an approximate symmetry.

Results obtained over the last few years (which are partly reviewed later in these notes) allow us to formulate a similar expectation from the general-relativity perspective. Loop quantum gravity and other discretized-spacetime quantum-gravity approaches appear to require a description of the familiar (continuous) Lorentz symmetry as an approximate symmetry, with departures governed by the Planck scale. And in the study of noncommutative spacetimes some Planck-scale departures from Lorentz symmetry appear to be inevitable.

From the particle-physics perspective there is instead no obvious reason to renounce to exact Lorentz symmetry, since Minkowski classical spacetime is an admissible background spacetime, and in classical Minkowski there cannot be any a priori obstruction for classical Lorentz symmetry. Still, a break up of Lorentz symmetry, in the sense of spontaneous symmetry breaking, is of course possible, and this possibility has been studied extensively over the last few years, especially in String Theory (see, e.g., [6] and references therein).

3.2 On the Fate of Poincaré Symmetry in Noncommutative Spacetimes

It is probably fair to say that we do not yet have a compelling proposed full solution of the quantum-gravity problem based on noncommutative geometry. But several heuristic arguments suggest that noncommutativity might be needed at the fundamental level, and most research on noncommutative spacetimes assumes that, if indeed noncommutativity is needed at the fundamental level in the most general setting, there should be some noncommutativity left over in the Minkowski limit of quantum gravity. Indeed, there has been much recent interest in flat noncommutative spacetimes, as possible quantum versions of Minkowski spacetime. Most of the work has focused on various parts of the two-matrix parameter space

$$[x_\mu, x_\nu] = i\frac{1}{E_p^2}Q_{\mu\nu} + i\frac{1}{E_p}C^\beta_{\mu\nu}x_\beta \ , \tag{2}$$

The assumption that the commutators of spacetime coordinates would depend on the coordinates at most linearly is adopted both for simplicity and because it

captures a very general intuition: assuming that the Planck scale governs noncommutativity (and therefore noncommutativity should disappear in the formal $E_p \to \infty$ limit), and assuming that the commutators do not involve singular $(1/x^n)$, terms one cannot write anything more general than (2).

Most authors consider two particular limits [73]: the "canonical noncommutative spacetimes", with $C^{\beta}_{\mu\nu} = 0$,

$$[x_\mu, x_\nu] = i\theta_{\mu\nu} \tag{3}$$

and the "Lie-algebra noncommutative spacetimes", with $Q_{\mu\nu} = 0$,

$$[x_\mu, x_\nu] = i\gamma^{\beta}_{\mu\nu} x_\beta \tag{4}$$

(I am adopting notation replacing $Q_{\mu\nu}/E_p^2 \to \theta_{\mu\nu}$ and $C^{\beta}_{\mu\nu}/E_p \to \gamma^{\beta}_{\mu\nu}$).

Results obtained over these past few years show [1–7, 74–77] that theories in these noncommutative spacetimes inevitably host some departures from Special Relativity (the only way to preserve Poincaré symmetry is the choice $\theta = 0 = \gamma$, i.e. the case in which there is no noncommutativity and one is back to the familiar classical commutative Minkowski spacetime).

And typically[13] (for most choices of the matrices θ and γ and a rather standard [6] field-theoretic description) Poincaré symmetry is broken, i.e. the departures from Special Relativity are such that the equivalence of inertial observers is lost.

As a simple way to see the departures from Poincaré symmetry one may consider canonical noncommutative spacetimes and look at the properties of the wave exponentials. The Fourier theory in canonical noncommutative spacetime is based [73] on simple wave exponentials $e^{ip^\mu x_\mu}$ and from the $[x_\mu, x_\nu] = i\theta_{\mu\nu}$ noncommutativity relations one finds that

$$e^{ip^\mu x_\mu} e^{ik^\nu x_\nu} = e^{-\frac{i}{2} p^\mu \theta_{\mu\nu} k^\nu} e^{i(p+k)^\mu x_\mu} \ , \tag{5}$$

i.e. the Fourier parameters p_μ and k_μ combine just as usual, with the only new ingredient of the overall phase factor that depends on $\theta_{\mu\nu}$. The fact that momenta combine in the usual way reflects the fact that the transformation rules for energy-momentum from one (inertial) observer to another are still the usual, undeformed, Poincaré transformation rules. However, the product of wave exponentials depends on $p^\mu \theta_{\mu\nu} k^\nu$, i.e. it depends on the "orientation" of the energy-momentum vectors p^μ and k^ν with respect to the $\theta_{\mu\nu}$ tensor. This is a first indication that in these canonical noncommutative spacetimes Poincaré symmetry is broken.[14] The $\theta_{\mu\nu}$ tensor plays the role of a background that can be used to introduce a preferred class of inertial observers. Different particles are

[13] There are some theories in some noncommutative spacetimes which are being considered as possible realizations of the "doubly-special-relativity" idea, but I shall comment on those in the next section.

[14] I should however warn the readers of the fact that recently a possible role of a 10-generator Poincaré-like Hopf algebra in certain formulations of theories in canonical noncommutative spacetimes has been considered by some authors (see, e.g., [78–80]).

affected by the presence of this background in different ways, leading to the emergence of different energy-momentum dispersion relations, as shown by the results [5,75,76] of the study of field theories in canonical noncommutative spacetimes. In particular one finds that birefringence characterizes the propagation of light (photons) in these spacetimes.

A condensed-matter analog of this type of spacetime-symmetry breaking is provided by the propagation of light in certain crystals: the rest frame of the crystal is a preferred frame for the description of the propagation, the crystal structure may be characterized in terms of a tensor, and indeed one encounters birefringence.

3.3 Spacetime and Poincaré Symmetry in String Theory

String Theory is the most mature quantum-gravity approach from the particle-physics perspective. As such it of course attempts to reproduce as much as possible the successes of quantum field theory, with gravity seen (to a large extent) simply as one more gauge interaction. Although the introduction of extended objects (strings, branes, ...) leads to some elements on novelty, in String Theory the core features of quantum gravity are essentially described in terms of graviton-like exchange in a background classical spacetime.

Indeed String Theory does not lead to spacetime quantization, at least in the sense that its background spacetime has been so far described as completely classical. However, this point is not fully settled: it has been shown that String Theory eventually leads to the emergence of a fundamental limitation on the localization of a spacetime event [81–84] and this might be in conflict with the assumption of a physically-meaningful classical background spacetime.

If eventually there will be a formulation of String Theory in a background spacetime that is truly quantum, it is likely that Poincaré symmetry symmetry will then not be an exact symmetry of the theory. If instead somehow a classical spacetime background can be meaningfully adopted, of course then there is no a priori reason to contemplate departures from Poincaré symmetry: classical Minkowski spacetime would naturally be an acceptable background, and a theory in the Minkowski background can be easily formulated in Poincaré symmetry-invariant manner.

Still, it is noteworthy that, even assuming that it makes sense to consider a classical background spacetime, the fate of Poincaré symmetry in String Theory is somewhat uncertain: it has been found that under appropriate conditions (a vacuum expectation value for certain tensor fields) Poincaré symmetry is broken. In these cases String Theory admits description (in the effective-theory sense) in terms of field theory in a noncommutative spacetime [6] with most of the studies focusing on the possibility that the emerging noncommutative spacetime is "canonical".

3.4 Spacetime and Poincaré Symmetry in Loop Quantum Gravity

Loop Quantum Gravity is the most mature approach to the quantum-gravity problem that originates from the general-relativity perspective. As for String Theory, it must be stressed that the understanding of this rich formalism is still very much "in progress". As presently understood, Loop Quantum Gravity predicts an inherently discretized spacetime [85]. There has been much discussion recently, prompted by the studies [9, 10], of the possibility that this discretization might lead to departures from ordinary Poincaré symmetry. Although there are cases in which a discretization is compatible with the presence of continuous classical symmetries [86–88], it is of course natural, when adopting a discretized spacetime, to put Poincaré symmetry under careful scrutiny. Arguments presented in [9–11], support the idea of broken Poincaré symmetry in Loop Quantum Gravity.

Moreover, recently Smolin, Starodubtsev and I proposed [13] (also see the follow-up study in [89]) a mechanism such that Loop Quantum Gravity would be described at the most fundamental level as a theory that in the flat-spacetime limit admits deformed Poincaré symmetry, in the sense of "doubly special relativity" (see next section). Our argument originates from the role that certain quantum symmetry groups have in the Loop-Quantum-Gravity description of spacetime with a cosmological constant, and observing that in the flat-spacetime limit (the limit of vanishing cosmological constant) these quantum groups might not contract to a classical Lie algebra, but rather contract to a quantum (Hopf) algebra.

4 More on the Quantum-Gravity-Inspired DSR Scenario

4.1 A "Doubly-Special" Relativity

As one can infer already from the observations reported in the previous section, the description of Planck-scale corrections to particle-physics processes will be a key aspect of the Minkowski limit of quantum gravity. In our current conceptual framework special relativity emerges in the Minkowski limit, where one deals with situations that allow the adoption of a Minkowski metric throughout, and one might wonder whether the Minkowski limit of quantum gravity could still be governed by special relativity. The issue will be of particular interest if quantum gravity admits a limit in which one can assume throughout a (expectation value of the) metric of Minkowski type, but some Planck-scale features of the fundamental description of spacetime (such as spacetime discreteness and/or spacetime noncommutativity) are still not completely negligible. In such a "nontrivial Minkowski limit" [90] essentially the role of the Planck scale in the description of gravitational interactions (expressing the gravitational constant G in terms of the Planck scale) can be ignored, but the possible role of the Planck scale in spacetime structure/kinematics is still significant. It is of course not obvious that the correct quantum gravity should admit such a nontrivial

Minkowski limit. With the little we presently know about the quantum-gravity problem we must be open to the possibility that the Minkowski limit would actually be trivial, i.e. that whenever the role of the Planck scale in the description of gravitational interactions can be neglected one should also neglect the role of the Planck scale in spacetime structure/kinematics. But it appears that the possibility of a nontrivial Minkowski limit is rather likely. Indeed, as illustrated by the observations reported in the previous section, for various approaches to the quantum-gravity problem evidence in support of a nontrivial Minkowski limit has emerged. The fate of Poincaré symmetry in such nontrivial Minkowski limits is of course a key issue both phenomenologically [9,20,21,23,36,51,91] and from a conceptual perspective. In the large number of studies produced between 1997 and 2000 (see, e.g., [9,10,20,21,36]) on the possibility of a nontrivial Minkowski limit it was always assumed that Poincaré (and in particular Lorentz) symmetry would be broken: the Galilei Relativity Principle would not hold with Planck-scale accuracy.

As already mentioned in Sect. 1, on the basis of an analogy with the process which led from Galilei/Newton Relativity to Einstein's Special Relativity, I argued in [14] that the Minkowski limit of quantum gravity might be characterized by a "doubly special relativity", a relativistic theory with two, rather than one, nontrivial relativistic invariants (the Planck scale in addition to the speed-of-light scale), but still fully compatible with the Galilei Relativity Principle.

The fact that Maxwell's equations involve a characteristic speed scale "c", and that speeds are not invariant according to Galilei Relativity, led at first the physics community to look for a "ether frame" in which to formulate Maxwell's equations. But in the end it was realized that the presence of the scale c in Maxwell's equations is not a manifestation of a preferred frame but rather a manifestation of the necessity to introduce a dimensionful (c-) deformation of the Galilei boosts. (FitzGerald-Lorentz-) Poincaré boosts are indeed a dimensionful deformation of Galilei boosts. Galilei Relativity was replaced by Special Relativity, but the transition preserves the equivalence of inertial observers. I made the hypothesis [14] that analogously the departures from Special Relativity induced by the role of the Planck length in certain quantum-spacetime pictures might not lead to a "broken" Poincaré symmetry (as always assumed in the 1997–2000 studies [9,10,20,21,36]), but should rather require yet another deformation of spacetime symmetry transformations, preserving the validity of the Galilei Relativity Principle.

At present this DSR proposal must be viewed as a physics scenario for which a fully satisfactory mathematical formalism is still being sought. The prototypical example of a quantum-gravity theory that would need the DSR idea is a quantum-gravity theory in which in the Minkowski limit one finds that the Planck (length) scale sets an observer-independent minimum allowed value of wavelength. And of course there are many other possible postulates that could be used to introduce a DSR Minkowski limit. While there is an abundance of possible physical principles that one might consider from a DSR perspective, there is still an intense debate on the choice of a formalism (or some formalisms)

which may be used to describe such DSR Minkowski limits. As I shall stress later on in these notes, one possibility that is attracting interest is the formalism of a certain noncommutative spacetime picture, and an associated Hopf-algebra [1, 2, 7, 49] description of spacetime symmetries. And (with or without an embedding within a Hopf-algebra structure) most authors are assuming that a DSR Minkowski limit would require a corresponding nonlinear realization of boost transformations. But all the formalisms considered so far are still lacking a compelling "interpretation theory" and/or provide an incomplete physical picture and/or appear to introduce some unplausible features.

In the remainder of this section I will try to summarize the preliminary results of this search of a formalism, placing emphasis on some robust results but also commenting on some key open issues.

4.2 An Illustrative Example (Discussed in Leading Order)

In proposing the idea of a DSR Minkowski limit, it felt necessary [14] to illustrate the idea with at least some formulas, and at least an illustrative example of new postulates (replacing the ones of Special Relativity) that would be of DSR type.

Concerning the postulates, one should start by noticing that any relativistic theory of the Minkowski limit will have to be based on Galilei Relativity Principle, as the first (and most important) postulate. This essentially states that the laws of physics are "the same" for every inertial observer. A second postulate would then specify one or more of these laws of physics: enough to characterize the relativistic theory, especially through the introduction of some nontrivial observer-independent scales.

Special Relativity describes the implications of the Relativity Principle for the case in which there is an observer-independent velocity scale c, setting the speed of all forms of light (independently of the light's wavelength, and the speed of the source emitting the light). Einstein's second postulate can be described in modern language as introducing a law of physics that links the energy and the (space-)momentum of photons: $E = c|p|$. In my first illustrative example of DSR-type second postulate I considered the law of physics $E^2 = c^2p^2 + f(E, p; c, \tilde{L}_p)$, for the relation between energy and (space-)momentum of photons, where f is some function which at first I will not fully specify but I will assume that its leading $\tilde{L}_p$ dependence be given by $f(E, p; L_p) \simeq \tilde{L}_p cEp^2$. As a second-postulate law, the law $E^2 = c^2p^2 + f(E, p; c, \tilde{L}_p)$ must be valid (with the same values of its parameters c and $\tilde{L}_p$) for every inertial observer, and can therefore be used to give physical meaning to two observer-independent scales: c and $\tilde{L}_p$. Each inertial observer can establish the value of c and $\tilde{L}_p$ (same values for all inertial observers) by determining the energy-momentum (dispersion) relation for photons. In the infrared limit the $E(p)$ relation will still be linear and the proportionality constant in the infrared limit is c. The second observer-independent scale $\tilde{L}_p$ will set the magnitude of departures from linearity as the energy of the photons increases.

In order to really have a new relativistic theory one should provide a complete description, consistent with the DSR postulates, of all the laws of Nature. Such

a programme has still not been brought to completion, not even for a single example of DSR postulates. It is however reassuring that at least one can verify that the illustrative example of postulates I am here considering (and some other examples which I will mention later in these notes) can be enforced very explicitly for some aspects of the kinematics of energy-momentum space. In particular, one may ask how should energy and momentum transform under Lorentz boosts and rotations in order to enforce $E^2 = c^2p^2 + f(E, p; c, \tilde{L}_p)$ as an observer-independent law. Let me first verify at first order in $\tilde{L}_p$ that this can be done. I shall therefore exhibit rotation and boost transformations such that $E^2 = c^2p^2 + \tilde{L}_p cEp^2$ is an invariant up to order-$\tilde{L}_p$ accuracy.

Actually, since the modified dispersion relation is still evidently invariant under classical (undeformed) space rotations, we can assume the rotation generators to be undeformed. The description of boosts should instead inevitably change, but it turns out [14] that it is possible[15] to simply introduce a deformed boost action. We can see this by first focusing on boosts in a chosen direction, say the z direction. Adopting the ansatz $B_z^{\tilde{L}_p} = i[cp_z + \tilde{L}_p\Delta_E]\partial/\partial E + i[E/c + \tilde{L}_p\Delta_{p_z}]\partial/\partial p_z$, one easily finds that the sought invariance translates into the requirement $2E\Delta_E - 2p_z\Delta_{p_z} = -2E^2p_z - p_z^3$. The simplest solutions are of the type $\Delta_E = 0,\ \Delta_{p_z} = E^2 + p_z^2/2$ and $\Delta_{p_z} = 0,\ \Delta_E = -Ep_z - p_z^3/(2E)$. Various arguments of simplicity [14] (including considerations involving combinations of boosts and rotations and the desire to have generators which would be well-behaved even off shell) lead me to adopt the option with $\Delta_E = 0,\ \Delta_{p_z} = E^2 + p_z^2/2$, so the new z-boost generator takes the form

$$B_z^{\tilde{L}_p} = icp_z\frac{\partial}{\partial E} + i[E/c - \tilde{L}_pE^2/c^2 - \tilde{L}_pp_z^2/2]\frac{\partial}{\partial p_z}\ . \tag{6}$$

It is useful to obtain explicit formulas for the finite z-boost transformations. The z-boost generator essentially describes infinitesimal transformations:

$$\frac{dE}{d\xi} = i[B_z^{\tilde{L}_p}, E] = -cp_z\ , \tag{7}$$

$$\frac{dp_z}{d\xi} = i[B_z^{\tilde{L}_p}, p_z] = -E/c - \tilde{L}_pE^2/c^2 - \tilde{L}_pp_z^2/2\ . \tag{8}$$

In spite of the richer structure of the new z-boost generator, the derivation of finite transformations from the structure of the generator of infinitesimal transformations is not significantly more complex than in the Lorentz case. With simple, but somewhat tedious, calculations one finds that

[15] While the simple deformation of boost action is clearly a possibility, it is not clear that this would be the only possibility for enforcing the postulates of the illustrative example. One ends up considering the deformation as a way to satisfy the postulates while keeping most of the physical picture of a boost, but it is conceivable that some more subtle mechanism (e.g., changing the "meaning" of boost action) could also be introduced.

$$E(\xi) = -c^{-1}\alpha e^{\xi} + c^{-1}\beta e^{-\xi} + \tilde{L}_p c^{-2}\frac{\alpha^2}{2}e^{2\xi} + \tilde{L}_p c^{-2}\frac{\beta^2}{2}e^{-2\xi} + \tilde{L}_p c^{-2}\alpha\beta \ , \quad (9)$$

$$p_z(\xi) = \alpha e^{\xi} + \beta e^{-\xi} - \tilde{L}_p \alpha^2 e^{2\xi} + \tilde{L}_p \beta^2 e^{-2\xi} \ . \quad (10)$$

where α and β depend on $E_0 \equiv [E]_{\xi=0}$ and $p_{z,0} \equiv [p_z]_{\xi=0}$:

$$\alpha = cp_{z,0}/2 - E_0/2 - \tilde{L}_p E_0 p_{z,0}/2 + \tilde{L}_p c p_{z,0}^2/4 \ , \quad (11)$$

$$\beta = cp_{z,0}/2 + E_0/2 - \tilde{L}_p E_0 p_{z,0}/2 - \tilde{L}_p c p_{z,0}^2/4 \ . \quad (12)$$

One can easily verify that indeed the $E(\xi)$ and $p_z(\xi)$ of (9) and (10) satisfy the relation $E(\xi)^2 = c^2 p_z^2(\xi) + \tilde{L}_p c E(\xi) p_z^2(\xi)$ with order-$\tilde{L}_p$ accuracy.

4.3 An Illustrative Example (All-Order Analysis)

It is actually rather easy to go beyond leading order in $\tilde{L}_p$ and to give a full 3+1D description of boosts. As an all order generalization of the illustrative example one may consider [53] a second postulate introducing the following dispersion relation:

$$\frac{e^{\tilde{L}_p E} + e^{-\tilde{L}_p E} - 2}{\tilde{L}_p^2} - \mathbf{p}^2 e^{\tilde{L}_p E} = \frac{e^{\tilde{L}_p m} + e^{-\tilde{L}_p m} - 2}{\tilde{L}_p^2} \ . \quad (13)$$

The case of photons (massless particles) is obtained through the requirement $m = 0$.

Then following some logical steps analogous to the ones described above for the leading-order analysis, one is led to the following form of the boost generators

$$N_k = ip_k\frac{\partial}{\partial E} + i\left(\frac{\tilde{L}_p}{2}\mathbf{p}^2 + \frac{1 - e^{-2\tilde{L}_p E}}{2\tilde{L}_p}\right)\frac{\partial}{\partial p_k} - i\tilde{L}_p p_k\left(p_j\frac{\partial}{\partial p_j}\right) \ . \quad (14)$$

In spite of the apparently virulent nonlinearities in this formula for boosts, finite boost transformation can once again be described explicitly. Let us see this focusing on the case of a particle which, for a given first observer, has four-momentum $(\omega^0, \mathbf{k}^0)$, and has four-momentum $(\omega, \mathbf{k})$ for a second observer in relative motion along the "1-axis", with boost/rapidity parameter ξ, with respect to the first observer. Using (14) one obtains the following differential equations for $(\omega(\xi), \mathbf{k}(\xi))$:

$$\frac{d}{d\xi}k_1(\xi) + \frac{\tilde{L}_p}{2}(k_1^2(\xi) - k_2^2(\xi) - k_3^2(\xi)) + \frac{e^{-2\tilde{L}_p\omega(\xi)} - 1}{2\tilde{L}_p} = 0 \quad (15)$$

$$\frac{d}{d\xi}\omega(\xi) - k_1(\xi) = 0 \quad (16)$$

$$\frac{d}{d\xi}k_2(\xi) + \tilde{L}_p k_1(\xi)k_2(\xi) = 0 \quad (17)$$

$$\frac{d}{d\xi}k_3(\xi) + \tilde{L}_p k_1(\xi)k_3(\xi) = 0 \quad (18)$$

Differentiating (15) and making use of the other equations one obtains a non-linear second-order equation for $k_1(\xi)$,

$$\frac{d^2}{d\xi^2}k_1(\xi) + 3\,\tilde{L}_p\,k_1(\xi)\frac{d}{d\xi}k_1(\xi) + \tilde{L}_p^2 k_1^3(\xi) - k_1(\xi) = 0\ . \tag{19}$$

whose solutions of are of the form

$$k_1(\xi) = -\frac{B}{\tilde{L}_p}\frac{\cosh(\xi+\beta)}{(1 - B\sinh(\xi+\beta))} \tag{20}$$

where B and β are integration constants (explicitly given, as functions of $\omega^0, \mathbf{k}^0$, in [53]).

Corresponding solutions for the other components of the four-momentum can be obtained by substituting (20) in (16)–(18). Therefore the general description of finite boost transformations can be given in close form.

A very efficient way (which also implicitly provides the integration constants B and β) to describe these deformed-boost transformation rules is through the amount of rapidity needed to take a particle from its rest frame to a frame in which its energy is ω (and its momentum is $k(\omega)$, which is fixed, once ω is known, using the dispersion relation and the direction of the boost):

$$\cosh(\xi) = \frac{e^{\tilde{L}_p\omega} - \cosh\left(\tilde{L}_p m\right)}{\sinh\left(\tilde{L}_p m\right)}\ , \quad \sinh(\xi) = \frac{\tilde{L}_p\,k\,e^{\tilde{L}_p\omega}}{\sinh\left(\tilde{L}_p m\right)}\ . \tag{21}$$

4.4 The Sign of $\tilde{L}_p$ and the Nature of the Nonlinearities in the Illustrative Example

In the previous subsection I gave an explicit description of finite boost transformations for an illustrative example of DSR second postulate. The formulas may at first suggest that $\tilde{L}_p$ (with dimensions of a length, and naturally to have modulus of the order of the Planck length scale) may be either positive or negative, but actually only positive $\tilde{L}_p$ is truly acceptable. This can be established already from an analysis of the structure of the equations (21): while for positive $\tilde{L}_p$ these equations are meaningful for any value of the rapidity, for negative $\tilde{L}_p$ these equations are meaningful only for relatively small values of rapidity, and for a critical finite value of rapidity a divergence of energy in encountered.

This is after all not so surprising, since the differential equations that govern the dependence of energy-momentum on rapidity are of a type that does not necessarily lead to the existence of global solutions $E(\xi)$, $p(\xi)$. The structure of the equations does not fulfill the standard Cauchy requirements for the existence of global solutions $E(\xi)$, $p(\xi)$. It is sufficient for us to discuss this issue considering a single boost (along a given direction). In this case one easily finds that the dependence of energy-momentum on rapidity is governed by the differential equations

$$\frac{dp(\xi)}{d\xi} = \frac{\tilde{L}_p}{2} p^2(\xi) + \frac{1 - e^{-2\tilde{L}_p E(\xi)}}{2\tilde{L}_p} \ , \tag{22}$$

$$\frac{dE(\xi)}{d\xi} = p(\xi) \ . \tag{23}$$

Within this context the Cauchy requirements can be compactly stated introducing the two-component function $Y(\xi) \equiv \{Y_1(\xi), Y_2(\xi)\} \equiv \{E(\xi), p(\xi)\}$, and using the notation $Y_l' \equiv dY_l/d\xi = F_l(Y)$ $(l \in \{1, 2\})$:

(i) F must be continuous;
(ii) for every $M \in \Re$ and for every $X \in \Re^2$ and $Z \in \Re^2$, such that $|X| \leq M$,$|Z| \leq M$, there must exist an $L_M \in \Re$, such that $\mid F(X) - F(Z) \mid \leq L_M \mid X - Z \mid$;
(iii) for every $X \in \Re^2$ there must exist $L_1 \in \Re$ and $L_2 \in \Re$ such that $\mid F(X) \mid \leq L_1 + L_2 \mid X \mid$.

(Of course, $|W|$ denotes $\sqrt{W_a^2 + W_b^2}$ for every $W \equiv \{W_a, W_b\} \in \Re^2$.)

The requirements (i) and (ii) are easily verified, but a possible problem for (iii) originates from the nonlinear structure of equation (22). The corresponding differential equations of ordinary special relativity ($\tilde{L}_p \to 0$ limit) are linear and automatically verify the Cauchy "sublinearity requirement" (iii). Instead the nonlinearity of the DSR differential equations imposes a detailed analysis. The system of equations (22)–(23) evidently satisfies the Cauchy requirements for existence and uniqueness of a local solution (in a neighborhood of a given value of ξ), but one is not a priori assured of the existence of a global solution.

A detailed analysis shows that for positive $\tilde{L}_p$ everything is ok: the Cauchy "sublinearity requirement" (iii) is satisfied (in spite of the nonlinearity of the equations) and therefore the existence of global solutions is assured. But for negative $\tilde{L}_p$ the Cauchy "sublinearity requirement" is not satisfied. The interested reader can straightforwardly (but somewhat tediously) verify that indeed for positive $\tilde{L}_p$ one can find two real numbers L_1 and L_2 with he property required in (iii). Instead for negative $\tilde{L}_p$ there is no pair of real numbers L_1 and L_2 such that requirement (iii) would be satisfied.

It is here sufficient to discuss a simplified proof, restricted to the case relevant for on-shell particles (whose energy and momentum satisfy the DSR-modified dispersion relation). Imposing the dispersion relation one can of course reduce the system of two differential equations to a single differential equation:

$$\frac{dE(\xi)}{d\xi} = \frac{1}{|\tilde{L}_p|} \sqrt{1 - 2\cosh(\tilde{L}_p m) e^{-\tilde{L}_p E(\xi)} + e^{-2\tilde{L}_p E(\xi)}} \ . \tag{24}$$

Here the Cauchy "sublinearity requirement" asks us to find a pair of real numbers L_1 and L_2 such that $dE(\xi)d\xi \leq L_1 + L_2 E(\xi)$ for every $E(\xi)$. Indeed for positive $\tilde{L}_p$ (where the exponentials in (24) are of the type $e^{-|\tilde{L}_p|E}$, and $e^{-|\tilde{L}_p|E} \leq 1$) one can find such pairs of real numbers. For example, the choice $L_1 = 1/\tilde{L}_p$ and $L_2 = 0$ is acceptable. Instead in the case of negative $\tilde{L}_p$ one finds that for any given pair of real numbers L_1 and L_2 there is always a value of E such

that, according to (24), $dE(\xi)/d\xi > L_1 + L_2 E(\xi)$. This is due to the fact that for negative $\tilde{L}_p$ the exponentials in (24) are of the type $e^{|\tilde{L}_p|E}$, and diverge exponentially for large E.

Another way to see that only positive $\tilde{L}_p$ is acceptable comes from a different type of analysis of the nonlinearities of the illustrative example of DSR setup which I am here considering. The generators of space rotations have not been modified, and taking into account the form (14) of the deformed boost generators one can easily verify that in this DSR setup the generators of rotations and boosts still close the Lorentz algebra. The action of the boost generators has been deformed nonlinearly but in such a way that the rotation-boost algebra is still satisfied. So clearly there must be [54, 55] a set of variables $\mathcal{E}, \mathcal{P}$, functions of mass, momentum and energy, such that on $\mathcal{E}$ and $\mathcal{P}$ the boost generators act in undeformed way. One easily finds that by posing

$$\frac{\mathcal{E}(E,m)}{\mu(m)} = \frac{e^{\tilde{L}_p E} - \cosh\left(\tilde{L}_p m\right)}{\sinh\left(\tilde{L}_p m\right)} \ , \quad \frac{\mathcal{P}(E,p,m)}{\mu(m)} = \frac{\tilde{L}_p \, p \, e^{\tilde{L}_p E}}{\sinh\left(\tilde{L}_p m\right)} \tag{25}$$

(where $\mu(m)$ is just a notation for $\mu^2 = \mathcal{E}^2 - \mathcal{P}^2$) and assuming that $\mathcal{E}$ and $\mathcal{P}$ transform according to standard (undeformed) boosts, then one obtains laws of transformation for E, p which are the ones discussed in the previous subsection for the DSR illustrative example. But while for positive $\tilde{L}_p$ the relations (25) are perfectly meaningful, for negative $\tilde{L}_p$ and values of $\mathcal{E}$ greater than a certain finite $\mathcal{E}^*$ there is no (real) solution E.

One therefore concludes that, while the infinitesimal transformations generated by the boosts (14) are perfectly ok both for positive and negative $\tilde{L}_p$, in considering finite boost transformations only positive $\tilde{L}_p$ is acceptable. The characterization of the nonlinearities discussed in this subsection also immediately allows us to conclude that, starting from the illustrative example I am discussing, one may consider an infinity of other analogous examples, by simply choosing other forms of the nonlinear relations (25). In particular, Magueijo and Smolin have argued [54] in favour of the following alternative choice:

$$\frac{\mathcal{E}(E,m)}{\mu(m)} = \frac{E(1-\tilde{L}_p m)}{m(1-\tilde{L}_p E)} \ , \quad \frac{\mathcal{P}(E,p,m)}{\mu(m)} = \frac{p(1-\tilde{L}_p m)}{m(1-\tilde{L}_p E)} \ . \tag{26}$$

4.5 Klein–Gordon and Dirac Equations within the Illustrative Example

While, as already stressed (and described in greater detail later on in these notes), the illustrative example I am considering is at present not developed to the point of describing a systematic DSR deformation of all the laws of physics, a significant number of results has been obtained. As an example let me comment in this subsection on how one can obtain from a modified second postulate, of the type considered within the illustrative example, some correspondingly deformed Klein–Gordon and Dirac equations in energy-momentum space.

For the DSR-deformed Klein–Gordon equation the exercise is rather elementary: the form of the Klein–Gordon equation in energy-momentum space reflects directly the form of the energy-momentum dispersion relation. Therefore one finds (see e.g. [92]) that the DSR-deformed Klein–Gordon energy-momentum-space wave function[16] $\tilde{\phi}(E, \mathbf{p})$ must satisfy

$$\left(\frac{2}{\tilde{L}_p^2} \left[\cosh(\tilde{L}_p E) - \cosh(\tilde{L}_p m) \right] - e^{\tilde{L}_p E} \mathbf{p}^2 \right) \tilde{\phi}(E, \mathbf{p}) = 0 \ . \tag{27}$$

It is just a little bit more laborious [93, 94] to derive the form of the DSR-deformed Dirac equation. It is convenient to start by recalling that the ordinary special-relativistic Dirac equation

$$(\gamma^\mu p_\mu - m)\, u(\mathbf{p}) = 0 \ , \tag{28}$$

where γ^μ are the familiar "γ matrices" and u contains right-handed and left-handed spinors,

$$u(\mathbf{p}) \equiv \begin{pmatrix} u_R(\mathbf{p}) \\ u_L(\mathbf{p}) \end{pmatrix} \ , \tag{29}$$

can be reformulated conveniently by describing spinors with space momentum $\mathbf{p}$ in terms of a pure Lorentz boost from the rest frame:

$$u_R(\mathbf{p}) = e^{\frac{1}{2}\boldsymbol{\sigma}\cdot\boldsymbol{\xi}} u_R(0) = \left(\cosh\left(\frac{\xi}{2}\right) + \boldsymbol{\sigma}\cdot\mathbf{n} \sinh\left(\frac{\xi}{2}\right) \right) u_R(0) \ , \tag{30}$$

and

$$u_L(\mathbf{p}) = e^{-\frac{1}{2}\boldsymbol{\sigma}\cdot\boldsymbol{\xi}} u_L(0) = \left(\cosh\left(\frac{\xi}{2}\right) - \boldsymbol{\sigma}\cdot\mathbf{n} \sinh\left(\frac{\xi}{2}\right) \right) u_L(0) \ , \tag{31}$$

where $\mathbf{n}$ is the unit vector in the direction of the boost (and therefore characterizes the direction of the space momentum of the particle) and on the right-hand sides of (30) and (31) the dependence on momentum is also present implicitly through the special-relativistic relations between the boost parameter ξ and energy E,

$$\cosh \xi = \frac{E}{m} \ , \tag{32}$$

and between energy and spatial momentum, $E^2 = \mathbf{p}^2 + m^2$ (for given mass m of the particle).

The Dirac equation may then be viewed as the requirement that left-handed and right-handed spinors cannot be distinguished at rest:

$$\begin{pmatrix} -I & F^+(\xi) \\ F^-(\xi) & -I \end{pmatrix} \begin{pmatrix} u_R(\mathbf{p}) \\ u_L(\mathbf{p}) \end{pmatrix} = 0 \ , \tag{33}$$

[16] This Klein–Gordon wave function in energy-momentum space should end up being (once the development of the formalism is completed) the (deformed ?) Fourier transform of of the Klein–Gordon wave function in spacetime.

where

$$F^{\pm}(\xi) = 2\left(\cosh^2\left(\frac{\xi}{2}\right) - \frac{1}{2} \pm \boldsymbol{\sigma}\cdot\mathbf{n}\sinh\left(\frac{\xi}{2}\right)\cosh\left(\frac{\xi}{2}\right)\right) \ . \tag{34}$$

The DSR deformation of the Dirac equation is most easily obtained using as starting point this derivation of the Dirac equation based on boosting rest-frame spinors. In the derivation of the DSR-deformed Dirac equation the only changes are introduced by the DSR deformations of the relations between energy and rapidity and between energy and momentum. Specifically for the DSR illustrative example which I have been considering one should make use of the relations

$$2\tilde{L}_p^{-2}\cosh(\tilde{L}_p E) - \mathbf{p}^2 e^{\tilde{L}_p E} = 2\tilde{L}_p^{-2}\cosh(\tilde{L}_p m) \ , \tag{35}$$

and

$$E(\xi) = m + \tilde{L}_p^{-1}\ln\left(1 - \sinh(\tilde{L}_p m)e^{-\tilde{L}_p m}(1-\cosh\xi)\right) \ . \tag{36}$$

From (36) one finds that the parameter ξ that characterizes boosts from the rest frame can be expressed as a function of the energy in the following way

$$\cosh\xi = \frac{e^{\tilde{L}_p E} - \cosh(\tilde{L}_p m)}{\sinh(\tilde{L}_p m)} \ . \tag{37}$$

Using these (35)–(37) the derivation of the Dirac equation leads to the result

$$\begin{pmatrix} -I & F^{+}_{\tilde{L}_p}(E,m) \\ F^{-}_{\tilde{L}_p}(E,m) & -I \end{pmatrix}\begin{pmatrix} u_R(\mathbf{p}) \\ u_L(\mathbf{p}) \end{pmatrix} = 0 \tag{38}$$

where

$$F^{\pm}_{\tilde{L}_p}(E,m) = \frac{e^{\tilde{L}_p E} - \cosh(\tilde{L}_p m) \pm \boldsymbol{\sigma}\cdot\mathbf{n}\left(2e^{\tilde{L}_p E}\left(\cosh(\tilde{L}_p E) - \cosh(\tilde{L}_p m)\right)\right)^{\frac{1}{2}}}{\sinh(\tilde{L}_p m)} \ . \tag{39}$$

Introducing

$$D_0^{\tilde{L}_p}(E,m) \equiv \frac{e^{\tilde{L}_p E} - \cosh(\tilde{L}_p m)}{\sinh(\tilde{L}_p m)} \tag{40}$$

and

$$D_i^{\tilde{L}_p}(E,m) \equiv \frac{n_i\left(2e^{\tilde{L}_p E}\left(\cosh(\tilde{L}_p E) - \cosh(\tilde{L}_p m)\right)\right)^{\frac{1}{2}}}{\sinh(\tilde{L}_p m)} \tag{41}$$

the DSR-deformed Dirac equation can be rewritten as

$$\left(\gamma^{\mu} D_{\mu}^{\tilde{L}_p}(E,m) - I\right) u(\mathbf{p}) = 0 \tag{42}$$

where again the γ^{μ} are the familiar "γ matrices".

The nature of this DSR deformation of the Dirac equation becomes more transparent by rewriting (41) taking into account the DSR dispersion relation (35):

$$D_i^{\tilde{L}_p}(\mathbf{p}, m) = \frac{e^{\tilde{L}_p E}}{\tilde{L}_p^{-1} \sinh(\tilde{L}_p m)} p_i \ . \tag{43}$$

In particular, as one should expect, in the limit $\tilde{L}_p \to 0$ one finds

$$D_i^{\tilde{L}_p}(E, m) \to \frac{E}{m} \ , \tag{44}$$

$$D_i^{\tilde{L}_p}(\mathbf{p}, m) \to \frac{p_i}{m} \ , \tag{45}$$

and the familiar special-relativistic Dirac equation is indeed obtained in the $\tilde{L}_p \to 0$ limit.

It is also easy to verify that the determinant of the matrix $(\gamma^\mu D_\mu^{\tilde{L}_p}(E, m) - I)$ vanishes, as necessary. In fact,

$$\begin{aligned} &\det\Big(\gamma^\mu D_\mu^{\tilde{L}_p}(E, m) - I\Big) \\ &\quad = \left(\sinh^2(\tilde{L}_p m) - \left(e^{\tilde{L}_p E} - \cosh(\tilde{L}_p m)\right)^2 + \frac{e^{2\tilde{L}_p E}}{\tilde{L}_p^{-2}} \mathbf{p}^2\right)^2 \\ &\quad = \left(\frac{e^{\tilde{L}_p E}}{\tilde{L}_p^{-2}} \left(-2\tilde{L}_p^{-2} \cosh(\tilde{L}_p E) + \mathbf{p}^2 e^{\tilde{L}_p E} + 2\tilde{L}_p^{-2} \cosh(\tilde{L}_p m)\right)\right)^2 = 0 \ , \end{aligned} \tag{46}$$

where the last equality on the right-hand side follows from the DSR dispersion relation.

4.6 Energy-Momentum Conservation within the Illustrative Example

Up to this point the discussion of the illustrative example has been confined to the one-particle sector. There is at least one aspect of multiparticle processes that relativistic kinematics really ought to describe: the law of conservation of energy-momentum in particle reactions. It is sufficient for me to focus here on the simple case of a process $a+b \to c+d$ (collision processes with incoming particles a and b and outgoing particles c and d) and to formulate the observations in an essentially one-space-dimensional context.[17] The special-relativistic kinematic requirements for such processes are $E_a + E_b - E_c - E_d = 0$ and $p_a + p_b - p_c - p_d = 0$. Using

[17] In three space dimensions one-space-dimensional kinematics is relevant for head-on a-b collisions producing c-d at threshold (when the kinematical conditions are only barely satisfied and therefore the particles produced do not have any energy available for momentum components in the directions orthogonal to the one of the head-on collision). Collisions at a particle-production threshold are after all the most interesting collisions, since they force us to insist on the fact that all inertial observers agree when the delicate kinematic balance of threshold production is realized.

the special-relativistic transformation rules, $dE_j/d\xi = -p_j$, $dp_j/d\xi = -E_j$, one immediately verifies that when the requirements are satisfied in one inertial frame they are also verified in all other inertial frames:

$$\frac{d(E_a + E_b - E_c - E_d)}{d\xi} = -p_a - p_b + p_c + p_d \ , \tag{47}$$

$$\frac{d(p_a + p_b - p_c - p_d)}{d\xi} = -E_a - E_b + E_c + E_d \ . \tag{48}$$

The requirements $E_a + E_b - E_c - E_d = 0$ and $p_a + p_b - p_c - p_d = 0$ cannot be imposed in the illustrative example of DSR framework which I am analyzing. This is easily understood already by looking at the leading-$\tilde{L}_p$-order formulas for the boost transformation rules $dE_j/d\xi = -cp_j$, $dp_j/d\xi = -E_j/c - \tilde{L}_p E_j^2/c^2 - \tilde{L}_p p_j^2/2$, with respect to which the requirements $E_a + E_b - E_c - E_d = 0$ and $p_a + p_b - p_c - p_d = 0$ are clearly not covariant.

One therefore must replace standard energy-momentum conservation by some new law. This is after all not surprising, and actually should be expected on the basis of the analogy with the transition from Galilei Relativity to Special Relativity. In Galilei Relativity velocities are composed linearly, as necessary in a theory without a privileged velocity scale, but in Special Relativity, with its observer-independent velocity scale, velocities cannot be combined linearly. Analogously it is clear that in DSR, with the availability of an observer-independent length (momentum) scale, the linear law of composition of momenta might have to be replaced. Of course, one can consider various ways to replace the linear law of energy-momentum conservation. A simple possibility is that one might still have equations that play exactly the same role as the one of linear conservation of energy-momentum, but introduce some nonlinearities in order to ensure covariance with respect to the deformed boost transformations. In particular, in leading-$\tilde{L}_p$-order one can easily verify that the requirements

$$E_a + E_b + \tilde{L}_p c p_a p_b - E_c - E_d - \tilde{L}_p c p_c p_d = 0 \ , \tag{49}$$

$$p_a + p_b + \tilde{L}_p (E_a p_b + E_b p_a)/c - p_c - p_d - \tilde{L}_p (E_c p_d + E_d p_c)/c = 0 \ , \tag{50}$$

are indeed covariant with respect to the deformed boosts of the illustrative example.

Deformed all-order-in-$\tilde{L}_p$ laws of conservation of energy-momentum that are consistent with the modified postulates can be most easily formulated exploiting the observation discussed in Subsect. 4.4, concerning the existence of a nonlinear map between some auxiliary variables $\mathcal{E}, \mathcal{P}$, transforming as components of an ordinary special-relativistic four-vector, and the DSR E, p energy-momentum. In particular, denoting with F the nonlinear map $E, p \to \mathcal{E}, \mathcal{P}$, one obtains [55] an all-order generalization of (49)–(50) by enforcing

$$0 = F^{-1} \left(\Sigma_j F(p_j) \right) \ , \tag{51}$$

which is by construction DSR covariant. [In (51), for notational convenience, I denoted with p the four-momentum entering the process for each particle, so for

the outgoing particles one should take a "negative" p (more precisely p will be the "antipode" [93] of the momentum of the outgoing particles).]

Clearly the requirement

$$0 = \Sigma_j F(p_j) \tag{52}$$

is also covariant, and there is actually an infinity of laws that one may consider, if the only requirement is covariance. Of course, in a given DSR theory some additional physical criteria [14, 55], besides covariance, will end up specifying the form of the energy-momentum "conservation" requirements.

One of the most interesting characteristics of these candidate conservation laws is the possibility to combine nonlinearly properties of different particles entering the process. The requirement (51) (and its leading-order approximation (49)–(50)) is an example of "mixing conservation law", in the sense that the nonlinear correction terms in (51) depend on properties of pairs of particles. The requirement (52) instead is an example of "nonmixing conservation", since it is obtained as a sum of terms each depending on the properties of a single particle.

4.7 Minimum Length and Minimum Wavelength in DSR

Several quantum-gravity papers adopt the hypothesis that the Planck length should set the minimum allowed value for wavelengths, but in most cases the statements remain void of physical meaning, since they do not address the relativistic implications. According to Special Relativity wavelengths are FitzGerald-Lorentz contracted and the laws of physics must be observer independent, so the introduction of a minimum wavelength must either be done specifying the class of observers for which the given value is applicable (Poicaré-symmetry breaking) or be done "*a la* doubly-special relativity", introducing the minimum wavelength as an observer-independent law at the price of enforcing a deformation of boosts such that the minimum wavelength does not get FitzGerald-Lorentz contracted.

Similarly, there is a large literature on a vaguely-stated hypothesis that the Planck length should set the absolute limit on the measurability of lengths, but these studies often do not even provide an explicit statement concerning whether this absolute limits apply to the measurement of proper lengths or to the measurement of the length in any frame. And when a limit for lengths in any frame is assumed, the authors often still (even now that there is some literature on DSR Minkowski limits) do not comment on the implications for Special Relativity.

The illustrative (however partial) example of DSR framework I have considered in the preceding subsections clearly may lead to an observer-independent minimum wavelength: the transformation laws are such that the momentum of a fundamental particle can never exceed a maximum value, set by $1/\tilde{L}_p$, and if one maintains the usual relation between momentum and wavelength ($p \sim 1/\lambda$) this

in turn implies a minimum value for wavelengths.[18] The nonlinear deformation of boost transformations is such that it saturates on wavelength of size $\tilde{L}_p$.

In spite of the fact that the illustrative example has at present only limited applicability we may also envision the emergence of an $\tilde{L}_p$ limit on lengths. To see this we can introduce the speed of photons as $dE(p)/dp$ (the standard "group-velocity law"), where $E(p)$ of course is the dependence of energy on momentum inferred from the dispersion relation, and perform a leading-$\tilde{L}_p$-order analaysis of a gedanken length-measurement procedure. Let us consider two observers each with its own (space-) ship moving in the same space direction, the z-axis, with different velocities (i.e. with some relative velocity), and let us mark "A" and "B" two z-axis points on one of the ships (the rest frame). The procedure of measurement of the distance AB is structured as a time-of-flight measurement: an ideal mirror is placed at B and the distance is measured as the half of the time needed by a first photon wave packet, centered at momentum p_0, sent from A toward B to be back at A (after reflection by the mirror). Timing is provided by a digital light-clock: another mirror is placed in a point "C" of the rest frame/ship, with the same z-axis coordinate of A at some distance AC, and a second identical wave packet, again centered at p_0, is bounced back and forth between A and C. The rest-frame observer will therefore measure AB as $AB' = v_\gamma(p_0) \cdot N \cdot \tau_0/2$, where N is the number of ticks done by the digital light-clock during the A→B→A journey of the first wave packet, τ_0 is the time interval corresponding to each tick of the light-clock ($\tau_0 = 2\,AC/v_\gamma(p_0)$), and $v_\gamma(p)$ is the speed of photons as obtained through $dE(p)/dp$.

The observer on the second (space-) ship, moving with velocity V with respect to the rest frame, will instead attribute to AB the value

$$AB'' = \frac{v_\gamma(p)^2 - V^2}{v_\gamma(p)} N \frac{\tau}{2} \ , \tag{53}$$

where p is related to p_0 through the DSR-deformed boost transformation and τ is the time interval which the second observer, moving with respect to the rest frame, attributes to each tick of the light-clock. It is easy to verify that τ is related to τ_0 by

$$\tau = \frac{v_\gamma(p_0)}{\sqrt{v_\gamma(p')^2 - V^2}} \tau_0 \ , \tag{54}$$

where p' is related to p_0 through the formula for boosts in a direction orthogonal to the one of motion of the photon. Combining (53) and (54) one easily obtains

$$AB'' = \frac{[v_\gamma(p)^2 - V^2] v_\gamma(p_0)}{v_\gamma(p)\sqrt{v_\gamma(p')^2 - V^2}} N \frac{\tau_0}{2} = \frac{v_\gamma(p)^2 - V^2}{v_\gamma(p)\sqrt{v_\gamma(p')^2 - V^2}} AB' \ . \tag{55}$$

[18] Clearly one can contemplate an alternative DSR scenario in which the deformed laws of transformation are introduced directly at the level of of wavelength/frequency, introducing the minimum-wavelength law, and then a maximum momentum for fundamental particles will arise if the usual relation $p \sim 1/\lambda$ between momentum and wavelength is enforced.

The implications of (55) for length contraction are in general quite complicated, but they are easily analyzed in both the small-V and the large-V limits (examined here of course in leading order in $\tilde{L}_p$). For small V and small momentum (large wavelength) of the probes (55) reproduces ordinary FitzGerald-Lorentz contraction. For large V (55) predicts that AB'' receives two most important contributions: the familiar FitzGerald-Lorentz term $(AB' \cdot \sqrt{c^2 - V^2})$ and a new term which is of order $\tilde{L}_p |p| AB' / \sqrt{c^2 - V^2}$. As V increases the ordinary FitzGerald-Lorentz contribution to AB'' decreases as usual, but the magnitude of the new correction term increases. Imposing $|p| > |\delta p| > 1/AB''$ (the probe wavelength must of course be shorter than the distance being measured) one arrives at the result $AB'' > \sqrt{c^2 - V^2} AB' + \tilde{L}_p AB' / (AB'' \sqrt{c^2 - V^2})$. This result clearly implies that $AB'' > \tilde{L}_p$ for all values of V.

4.8 Some Characteristic Elements of DSR Phenomenology

Over the last decade there has been a sharp surge in the amount of research devoted to the phenomenology of Planck-scale departures from Special Relativity. At first of course these works focused on broken-symmetry scenarios, and even after the introduction of doubly-special relativity only a relatively small fraction of the overall number of studies devoted to the phenomenology of Planck-scale departures from Special Relativity considers the symmetry-deformation scenario. This is perhaps due to the fact that we still do not have examples of DSR frameworks which are developed to the point of providing a complete theory: in most cases one has some elements of a DSR kinematics and a very tentative description of some aspects of dynamics, but no comprehensive description of the laws of Nature.

In this subsection I want to stress that, in spite of the present limited development of DSR frameworks, through the analysis of the general conceptual structure of the DSR proposal, and using the intuition emerging from the analysis of some illustrative examples (such as the illustrative example considered in the preceding subsections), one can establish at least a few characteristic elements of DSR phenomenology. In particular, it is easy to establish that some DSR predictions are completely different from the corresponding predictions of broken-symmetry scenarios.

Perhaps the most striking observation comes from comparing the expectations for photon stability that one obtains in DSR and in broken-symmetry scenarios. It is well established [51, 95–97] that when Lorentz symmetry is broken at the Planck scale it is rather natural to find photon decay. Let us for example analyze the process $\gamma \to e^+ e^-$ using the same dispersion relation I have so far considered from a DSR perspective, but assuming (as in a popular broken-Lorentz-symmetry scenario [21, 23, 51, 95]) that instead Lorentz symmetry is broken and the law of energy-momentum conservation is not modified. In this scenario one easily finds a relation between the energy E_γ of the incoming photon, the opening angle θ between the outgoing electron-positron pair, and the energy E_+ of the outgoing positron (of course the energy of the outgoing electron is simply given by $E_\gamma - E_+$). For the region of phase space with

$m_e \ll E_\gamma \ll 1/\tilde{L}_p$ one obtains

$$\cos(\theta) \simeq \frac{E_+(E_\gamma - E_+) + m_e^2 - \eta \tilde{L}_p E_\gamma E_+ (E_\gamma - E_+)}{E_+(E_\gamma - E_+)} \ , \tag{56}$$

where m_e is the electron mass and $\eta \equiv \tilde{L}_p/|\tilde{L}_p|$ (η is the sign of $\tilde{L}_p$).

The fact that for $|\tilde{L}_p| \to 0$ (56) would require $\cos(\theta) > 1$ reflects the fact that if Lorentz symmetry is preserved the process $\gamma \to e^+e^-$ is kinematically forbidden. For negative $\tilde{L}_p$ ($\eta = -1$) the process is still forbidden, but for positive $\tilde{L}_p$ ($\eta = 1$) high-energy photons can decay into an electron-positron pair. In fact, for positive $\tilde{L}_p$ and $E_\gamma \gg (m_e^2 \tilde{L}_p)^{1/3}$ one finds regions of phase space where $\cos(\theta) < 1$.

The fact that one finds that a certain particle decay can occur only at energies higher than a certain minimum decay energy ($E_\gamma \gg (m_e^2 E_p/|\eta|)^{1/3}$) is a very explicit manifestation of the break down of Lorentz symmetry. A given photon will have high energy for some inertial observers and low energy for other inertial observers. And of course one would like to exclude the possibility[19] that the decay be allowed according to some observers and disallowed according to others. So clearly such a picture requires a preferred frame: the energy of the particle should be measured in the preferred frame and the decay is allowed if the energy of the particle in the preferred frame exceeds a certain given value.

For the particle-decay picture of Lorentz-symmetry-breaking scenarios the existence of a preferred class of inertial observers is therefore a prerequisite. And even without any calculations we can conclude that there cannot be any such mechanism in a truly consistent DSR scenario, which by definition must preserve the equivalence of inertial frames. If one day we had data showing that a certain particle decay can occur at high energies but cannot occur at low energies, this would be sufficient to rule out completely the DSR idea. At least in this respect the DSR idea is falsifiable.

This result originates simply from the required equivalence of inertial frames, and therefore it is independent of the specific choice of second postulate adopted in a given DSR framework. One can easily verify explicitly that the process $\gamma \to e^+e^-$ is not allowed in the illustrative example of DSR framework considered in the preceding subsections. We must simply combine the DSR-deformed dispersion relation with the DSR-deformed energy-momentum conservation law which in this case takes the form

$$E_\gamma \simeq E_+ + E_- - \lambda \mathbf{p}_+ \cdot \mathbf{p}_- \ , \quad \mathbf{p}_\gamma \simeq \mathbf{p}_+ + \mathbf{p}_- - \lambda E_+ \mathbf{p}_- - \lambda E_- \mathbf{p}_+ \ . \tag{57}$$

Then, considering again the region of phase space with $m_e \ll E_\gamma \ll E_p \sim 1/\tilde{L}_p$, one easily finds that the relation between E_γ, the opening angle θ, and E_+ must take the form

[19] I am assuming the "objectivity of particle-production processes" [14]. If according to one observer the "in state" (at time "$-\infty$") is a photon and the "out state" (at time "$+\infty$") is composed of an electron-positron pair, then all other observers should agree.

$$\cos(\theta) \simeq \frac{2E_+(E_\gamma - E_+) + 2m_e^2 + \tilde{L}_p E_\gamma E_+(E_\gamma - E_+)}{2E_+(E_\gamma - E_+) + \tilde{L}_p E_\gamma E_+(E_\gamma - E_+)} \ , \tag{58}$$

We see clearly from (58) that the photon is stable in the illustrative example of DSR framework I have been considering, as it should be in any genuine DSR framework.

The case of photon stability illustrates most powerfully that there should be some striking differences between DSR phenomenology and the phenomenology of scenarios with broken Lorentz symmetry. It is instructive to consider the differences between DSR and broken-symmetry scenarios also in the popular subject [20–23,51,95–97] of the implications of departures from Special Relativity in the analysis of certain types of energy thresholds for particle-production processes. It is here sufficient to consider the example of collisions between a soft photon of energy ϵ and a high-energy photon of energy E that creates an electron-positron pair: $\gamma\gamma \to e^+e^-$. For given soft-photon energy ϵ, the process is allowed only if E is greater than a certain threshold energy E_{th} which depends on ϵ and m_e^2. In the broken-Lorentz-symmetry scenario with the same dispersion relation of my illustrative example of DSR, but with undeformed energy-momentum conservation, one easily obtains (assuming $\epsilon \ll m_e \ll E_{th} \ll 1/|\tilde{L}_p|$)

$$E_{th}\epsilon + \tilde{L}_p \frac{E_{th}^3}{8} \simeq m_e^2 \ . \tag{59}$$

The special-relativistic result $E_{th} = m_e^2/\epsilon$ corresponds of course to the $\tilde{L}_p \to 0$ limit of (59). For $\tilde{L}_p \neq 0$ the correction can be safely neglected as long as $\epsilon/E_{th} > |\tilde{L}_p|E_{th}$. But eventually, for sufficiently small values of ϵ and correspondingly large values of E_{th}, the Planck-scale correction cannot be ignored.

These "threshold anomalies" [23] are also allowed in principle in a DSR framework, but (for given choice of the energy-momentum dispersion relation) one typically finds [98] that the DSR threshold anomalies are much smaller than the threshold anomalies found in broken-Lorentz-symmetry scenarios. One can easily verify this by analyzing again the process $\gamma\gamma \to e^+e^-$ within the DSR illustrative example which I have been considering. In this case the threshold calculation will be also affected, besides the modification of the dispersion relation, by the correlated modification of the law of energy-momentum conservation, which for $\gamma\gamma \to e^+e^-$ takes the form

$$E + \epsilon - \tilde{L}_p \mathbf{P}\cdot\mathbf{p} \simeq E_+ + E_- - \tilde{L}_p \mathbf{p}_+\cdot\mathbf{p}_- \ , \tag{60}$$

$$\mathbf{P} + \mathbf{p} - \tilde{L}_p E\mathbf{p} - \tilde{L}_p \epsilon \mathbf{P} \simeq \mathbf{p}_+ + \mathbf{p}_- - \tilde{L}_p E_+ \mathbf{p}_- - \tilde{L}_p E_- \mathbf{p}_+ \tag{61}$$

where $\mathbf{P}$ is the momentum of the photon of energy E and $\mathbf{p}$ is the momentum of the photon of energy ϵ.

The presence of correction terms both in the dispersion relation and in the energy-momentum-conservation law (with the ratio of the coefficients fixed by

the requirement of equivalence of inertial frames) leads to rather large cancellations in the threshold formula. Assuming again that $\epsilon \ll m_e \ll E_{th} \ll E_p$ one ends up finding

$$E_{th} \simeq \frac{m_e^2}{\epsilon} \ , \tag{62}$$

i.e. (for $\epsilon \ll m_e \ll E_{th} \ll E_p$) one ends up with the same result as in the special-relativistic case. If, rather than working within the approximations allowed by the hierarchy $\epsilon \ll m_e \ll E_{th} \ll E_p$, one considers the exact DSR threshold formula, one finds a result which is actually different from the special-relativistic one. There are "threshold anomalies" in the DSR scenario, but they are typically smaller[20] than in broken-Lorentz-symmetry scenarios.

4.9 A DSR Minkowski Limit Described by the κ-Poincaré Hopf Algebra?

While the physical concept of a DSR Minkowski limit is well defined, and has been well received by a sizeable community, the DSR literature has still not identified a fully acceptable mathematical formalism for the description of such a Minkowski limit. The nonlinear realizations of the Lorentz boosts on which my illustrative example is based may or may not be needed/useful: they provide an "ok description" of certain aspects of the kinematics of energy-momentum space, but nobody has been able to obtain from them a comprehensive theoretical framework free from inconsistencies and from troubles at the level of the interpretation of formulas. Most authors would agree that with the information presently available to us the formalism of the κ-Poincaré Hopf algebra [1, 2, 100] should be considered as the most promising opportunity for DSR research, even though several grey areas remain.

The properties of the κ-Poincaré Hopf algebra are most easily understood [7] considering the dual [1–3, 7, 74, 77] κ-Minkowski noncommutative spacetime $(l, m = 1, 2, 3)$

$$[x_m, t] = \frac{i}{\kappa} x_m \ , \qquad [x_m, x_l] = 0 \ . \tag{63}$$

κ-Minkowski is a Lie-algebra spacetime (see Sect. 3) that clearly enjoys classical space-rotation symmetry. From a DSR perspective it is particularly noteworthy that in κ-Minkowski boost transformations are necessarily modified. A first hint of this comes from the necessity of a deformed law of composition of wave exponentials, encoded in the so-called "coproduct" [1, 2]. For a given ordering convention for wave exponentials, such as

$$e^{ik^m x_m} e^{ik^0 x_0} \ , \tag{64}$$

one finds

[20] There are however some *ad hoc* ways, based on appropriately tailored dispersion relations, for obtaining large threshold anomalies in certain types of DSR scenarios [99].

$$e^{ip^m x_m} e^{ip^0 x_0} e^{ik^m x_m} e^{ik^0 x_0} = e^{i(p\dot{+}k)^m x_m} e^{i(p^0+k^0)x_0} \ , \tag{65}$$

where the notation "$\dot{+}$" here introduced reflects the behaviour of the mentioned "coproduct"

$$p_\mu \dot{+} k_\mu \equiv \delta_{\mu,0}(p_0 + k_0) + (1 - \delta_{\mu,0})(p_\mu + e^{\tilde{L}_p p_0} k_\mu) \ . \tag{66}$$

The κ-Poincaré Hopf algebra contains the generators of rotations, boosts and translations for κ-Minkowski. While the action of translations and rotations is classical (underformed) the requirement of consistency of the emerging Hopf algebra leads one to consider boosts that have a nonlinearly-deformed action. More explicitly: for translations P_μ and rotations M^j, once a "Weyl map" Ω is introduced,[21] one can insist on

$$P_\mu \Omega(e^{ikx}) = \Omega(P_\mu e^{ikx}) = k_\mu \Omega(e^{ikx}) \ , \tag{67}$$

$$M_j \Omega(e^{ikx}) = \Omega(M_j e^{ikx}) = \Omega(-i\epsilon_{jkl} x_k \partial_l e^{ikx}) \ , \tag{68}$$

whereas for boosts the analogous requirement would lead to an inconsistent description of symmetry transformations, since then some of the co-commutators (coproducts) would not close on the algebra [7]. The precise form of the deformed boost action (and, as a result, the form of the commutators between boost generators and other generators) depends on the choice of ordering convention for the wave exponentials (the choice of "basis" for the κ-Poincaré Hopf algebra). The ordering convention adopted in (64) leads to a deformed boost action that matches perfectly the one considered in my illustrative DSR example (14).

And other results which one can independently obtain starting from a given choice of DSR second postulate can be easily reproduced within the κ-Poincaré/κ-Minkowski framework. In particular, the form of the energy-momentum-space Klein–Gordon and Dirac equations within the DSR illustrative example which I considered above can be obtained as natural Klein–Gordon and Dirac equations for κ-Minkowski, upon appropriate choice of ordering convention. This has been recently verified explicitly [93] by essentially revisiting from a DSR perspective some previous κ-Poincaré/κ-Minkowski results [100–104].

All this is of course of some encouragement for the idea of using κ-Poincaré mathematics in the development of DSR physical theories, and further encouragement comes from the recent success in overcoming some of the apparent obstructions, but some key grey areas for the applicability of κ-Poincaré mathematics in DSR research remain. In the remainder of this subsection I intend to describe some of the apparent obstructions which were overcome, and some of the residual grey areas.

[21] A Weyl map connects a given function in the noncommutative Minkowski with a corresponding function in commutative Minkowski, and is not unique. It is sufficient to specify a Weyl map on the complex exponentials and extend it to the generic function $f(x)$, whose Fourier transform is $\tilde{f}(p) = \frac{1}{(2\pi)^4} \int f(x) e^{-ipx} \mathrm{d}^4 x$, by linearity $\Omega(f) = \int \tilde{f}(p)\, \Omega(e^{ipx})\, \mathrm{d}^4 p$.

Only a Quasigroup?

The discovery of the κ-Poincaré Hopf algebra [100] was not motivated by the desire of addressing some specific physics issues, but rather as a solution to a long-standing problem for the "quantum groups" literature: "quantum" ("q-deformed") versions of a large variety of algebras were being found, mainly using the Drinfeld-Jimbo scheme [105], but for the Poincaré algebra the Drinfeld-Jimbo scheme is not applicable. Lukierski, Ruegg, Nowicki and collaborators [100] realized that from the q-deformation of the de-Sitter algebra one could obtain, by appropriate Inonu-Wigner contraction, a quantum (Hopf) Poincaré algebra, the κ-Poincaré algebra, which however was not a q-deformed version of Poincaré (a dimensionless deformation) but rather a dimensionful deformation. Many "basis" for this Hopf algebra have been considered, but before the DSR proposal no genuine group of symmetry transformations had been discussed. Indeed some studies [106] concerning the exponentiation of Lorentz-like κ-Poincaré algebra generators had led to disappointing results: rather than the familiar group of finite symmetry transformations one would obtain by exponentiation only a "quasigroup" (in the sense of Batalin [107]). Even though these results had been obtained [106] working with a specific κ-Poincaré basis, they had led to the expectation that in general κ-Poincaré could not give rise to a genuine group of finite symmetry transformations. A result of DSR research which may be viewed as independently also valuable from a pure κ-Poincaré perspective is the one of [53], which identifies the first example of κ-Poincaré basis that leads, by exponentiation of the generators, to a genuine group of finite transformations, and analyses this transformations from a DSR perspective. This result provided much encouragement for applications of κ-Poincaré mathematics in DSR research: with a consistent algebra one can at best describe infinitesimal symmetry transformations, something which is of course not sufficient for a relativistic theory, whereas the finite transformations introduced in [53] essentially established that κ-Poincaré mathematics could at least be used to obtain a description of one-particle energy-momentum which is consistent with the DSR criteria.

A Sign Ambiguity?

The description of finite transformations obtained in [53], especially as later refined in [108], also addresses another potential issue for the possibility of using κ-Poincaré mathematics in the description of Planck-scale physics. The κ-Poincaré Hopf algebra has a free dimensionful parameter with arbitrary overall sign. While the magnitude of this parameter can be fixed "physically" by assuming that it should be of the order of the Planck scale, the residual sign ambiguity could be perceived as disappointing. However, while indeed the κ-Poincaré Hopf algebra is consistent for either choice of sign, the requirement of using this mathematics for the construction of a DSR framework, where the consistency of finite transformations is required (whereas the consistency of a Hopf algebra concerns only

the properties of the generators), allows, as discussed in Subsect. 4.4, to exclude one of the sign choices.

No Clear Physical Meaning?

The fact that the DSR perspective on κ-Poincaré led to the first analyses of finite κ-Poincaré transformations has also provided an opportunity for an intuitive physical characterization of the deformation parameter. In the " κ-Poincaré basis" that is consistent with the illustrative DSR example I have been considering (and in some other related bases) the deformed boost transformations of momenta of fundamental particles are such that they saturate at a maximum (minimum) $1/\tilde{L}_p$ value of space momentum (wavelength). Therefore the parameter that characterizes the definining algebraic relations for generators of the κ-Poincaré Hopf algebra, once the generators are exponentiated and used to describe finite transformations, may be used to introduce a minimum wavelength.

Also, while originally the motivation for obtaining the κ-Poincaré Hopf algebra came just from the desire of extending the applicability of the quantum-group research programme, the DSR literature has provided arguments for seeing the emergence of at least some κ-Poincaré properties within certain Minkowski limits of quantum-gravity theories. Most notably, it was noticed [13] that, according to a popular perspective on the Loop Quantum Gravity research programme, the deSitter limit of quantum gravity should be governed by a q-deformed deSitter algebra, and that it is rather plausible that then the Inonu-Wigner contraction to the Minkowski limit might give rise (following the technical steps of [100]) to a description based on the κ-Poincaré algebra. This would finally legitimize the assumption that the deformation parameter be identified with the Planck scale.

Inconsistent Laws of Energy-Momentum Conservation?

Up to this point I have mentioned some aspects of the κ-Poincaré mathematics which at first appeared to represent obstructions for its use in DSR research, but were eventually successfully handled in the DSR literature. But it is now time for me to mention some key residual open issues for the applicability of κ-Poincaré mathematics in DSR research. It is legitimate to hope that also these issues may eventually be handled in a satisfactory way from the DSR perspective, but they have been confronting the DSR community for some time, without significant improvements, so one must acknowledge the risk that these aspects of κ-Poincaré mathematics might force us to renounce to the use of κ-Poincaré in the description of DSR Minkowski limits.

Among these open issues very important is the one that concerns the description of energy-momentum conservation in particle-collision/particle-production processes. The idea of relativistic transformations between inertial observers was very far from the objectives of pre-DSR work on κ-Poincaré Hopf algebras (as mentioned, their motivation came from within the subject of quantum groups), and it is therefore not surprising that the recipe adopted in the κ-Poincaré

literature (see, e.g., [74] and references therein) for the description of energy-momentum conservation, based on the coproduct sum (66) and an action of boosts on the composed momenta induced by the action on each of the momenta entering the composition, is not relativistically acceptable. In practice from this κ-Poincaré perspective one assumes that, for example, for a particle-producing collision process $a + b \to c + d$ the "energy-momentum-conservation" condition would be $(p_a \dot{+} p_b)^\mu = (p_c \dot{+} p_d)^\mu$. Since no specific frame is mentioned in the relevant studies [74] one should think that this be valid in any frame, but from a relativistic perspective this is inconsistent with the description of the action of boosts on composite momenta obtained [74, 109] by simply acting with the boosts on the momenta of each particle ($N(p \dot{+} k) = [N(p)] \dot{+} [N(k)]$). It is easy to verify [88] that the condition $(p_a \dot{+} p_b)^\mu = (p_c \dot{+} p_d)^\mu$ is incompatible with the condition $(p'_a \dot{+} p'_b)^\mu = (p'_c \dot{+} p'_d)^\mu$ (where p_j and p'_j indicate the momentum of the j-th particle in two frames). If $(p_a \dot{+} p_b)^\mu = (p_c \dot{+} p_d)^\mu$ holds in one frame then necessarily $(p'_a \dot{+} p'_b)^\mu \neq (p'_c \dot{+} p'_d)^\mu$ in some other frames. The law $(p_a \dot{+} p_b)^\mu = (p_c \dot{+} p_d)^\mu$ is still advocated even in the post-DSR κ-Poincaré literature (see, e.g., [109]), but clearly it does not make sense relativistically.

If one should obtain a DSR theory using κ-Poincaré mathematics, then there should be some other natural way to describe energy-momentum conservation in terms of some structures within the κ-Poincaré Hopf algebra, but this has not yet been found. One possibility that one may consider [110] requires giving up the objectivity, in collision processes, of the momenta of individual particles entering the process: one could attribute to a four-particle system (for example, two particles entering the process and two "outgoing" particles) a total momentum P_μ^{tot} and assume that this total momentum transforms *a la* κ-Poincaré. Then the requirement $P_\mu^{tot} = 0$ could be upheld in all frames, but the rule for obtaining P_μ^{tot} as coproduct sum of momenta would then give to the particles participating in the process momenta which are not "relativistically objective" (p'_j not obtained from p_j via a boost). This would of course be only acceptable if the individual momenta of the particles participating in a process would not be good observables at the Planck-scale level.

How Do We "Choose" a Basis for Energy-Momentum?

Perhaps even more significant is the problem of the "choice of basis" for κ-Poincaré. From a mathematics perspective one can satisfactorily characterize a Hopf algebra by giving commutators and cocommutators (coproducts and the associated antipode and counit [2]) for a set of generators, but any nonlinear redefinition of the generators ($A_i \to B_i(A_j)$), although it introduces rather profound modifications in the structure of commutators and cocommutators, provides an equally good way to characterize the Hopf algebra. As mentioned, for the case of the κ-Poincaré Hopf algebra some changes of basis are very intuitively described in terms of the dual κ-Minkowski spacetime: for example one basis provides a natural description [7] of the space of functions of κ-Minkowski coordinates if functions are conventionally written with the time-coordinate dependence to the right (e.g., xy^2t), and another basis provides a natural description of the space

of functions of κ-Minkowski coordinates if functions are conventionally written treating time and space symmetrically (e.g., $xy^2t + txy^2$).

So far so good: same space of functions, which can be equivalently described using different conventions, associated with the same Hopf algebra, which is correspondingly written in different "basis". But the present description of energy-momentum within κ-Poincaré [1–3, 7, 74] is strongly basis dependent! The same Lagrangian theory in κ-Minkowski leads to different [7] concepts of energy-momentum depending on the choice of conventions! Clearly we have not understood how to describe energy-momentum in the κ-Minkowski/κ-Poincaré framework. One possible way out is the eventual discovery that only one of the basis allows the standard description of energy-momentum. But it is also possible that "choice of basis problem" actually turns out to be a problem of description of energy-momentum: perhaps one can have a good picture in any basis upon appropriate improved description of energy-momentum.

4.10 Other Key Open Issues

Since the majority of DSR publications implicitly or explicitly explore the possibility of using κ-Poincaré mathematics in the description of DSR Minkowski limits, all the problems mentioned in the preceding subsection are also to be viewed as problems for DSR research. There are a few other open problems whose study can be motivated intrinsically from the concept of a DSR Minkowski limit, rather than through the hope of relying on κ-Poincaré mathematics. In this subsection I want to briefly mention some of these other open problems.

How should One Describe Macroscopic Bodies in DSR?

One key issue for the DSR research programme concerns the description of macroscopic bodies. As stressed above, the availability of an observer-independent length/momentum scale provides an opportunity for nonlinearities in the energy-momentum sector, and these nonlinearities, even when typically small for fundamental particles, may add up to implausibly large new effects for macroscopic bodies. This is in particular a key concern for the much-studied possibility of a DSR Minkowski limit characterized by Planck-scale modifications of the energy-momentum dispersion relation for fundamental particles. While such Planck-scale modifications of the dispersion relation are clearly admissible for microscopic particles (they are always observed with momentum/energy much smaller that the Planck scale), they can easily lead to huge effects for a macroscopic body, whose energy will be much larger than the Planck scale.

This problem is automatically absent for some forms of the dispersion relation (see, e.g., [99]), and even when it could potentially be present it has been observed (already in the first papers on the DSR idea [14]) that the same nonlinearities that are producing the difficulty also appear to provide ways to overcome it. Essentially in these cases there must be some mechanism such that macroscopic bodies are governed by dispersion relations that are different from the ones applicable to microscopic bodies.

Another possibility [98] is that also the relation between momentum and wavelength be affected by the Planck scale. In that case one could have a nonlinear wavelength/frequency relation, while preserving the linearity of the energy-momentum dispersion relation. If the nonlinearities are confined to the wavelength/frequency relation the description of macroscopic bodies may be less challenging [98].

What is the DSR Observer?

An issue which is possibly related to this "macroscopic-body problem" is the one that concerns the description of DSR observers. For example, it is not implausible that the DSR observer be associated with a macroscopic system, to which the Planck-scale deformation does not apply. As long as the description of macroscopic bodies is not settled, we might still have an incomplete understanding of DSR observers.

At least for DSR proposals advocating a modified dispersion relation, it also appears (as first stressed in [98], and recently rediscovered in [111]) that one should think of these observers as equipped with a large variety of probes. Since the speed of photons can be wavelength/energy dependent, some of the perceptions of the observer should perhaps depend on the type of probes the observer uses in a given context. For example, in a context probed with high-energy probes the observer might experience a different type of time dilatation and length contraction. And perhaps, rather than thinking of a single observer with different types of probes, we should think of different types of observers, characterized by the type of probes they use.

At the merely technical level these issues concerning the DSR observer are relevant for establishing the relation between rapidity and relative velocity among observers. Is the special-relativistic relation still valid?

What about Causality?

Of course, another key issue for DSR theories will be the one of causality: should the introduction of the second observer-independent scale affect causality? For example, if indeed, as assumed in several studies, the second observer-independent scale is introduced in such a way that ultra-high-energy photons would travel faster than the low-energy photons which we ordinarily observe/study (whose speed is well established to be given by the "speed-of-light scale" c) the implications for causality should be profound. Actually, in a DSR theory one might be able to accommodate a formal limit in which a single particle has infinite energy and infinite speed.[22]

[22] Of course, we cannot even contemplate a particle with infinite energy (we can at best, very optimistically of course, contemplate the possibility to put all the energy of the Universe in a single particle), but nevertheless the given relativistic theory might predict a dependence of speed on energy which diverges in the infinite-energy limit, and in that case (in spite of the formal nature of the infinite-energy limit) there would be important implications for causality.

This might even have some welcome consequences: for example, a few authors have used (see, e.g., [57]) some preliminary intuition about the "new DSR causality" to construct cosmological models based on DSR that would not require inflation. These are the proposals, already mentioned above, which provide a new variant of "time-varying-speed-of-light cosmology".

But several aspects of causality might require a careful analysis in any given proposed DSR framework. Would the observers which have ultra-high-energy photons at their disposal be able to introduce a concept of time that is (to good approximation) absolute? The dilatation of the muon lifetime due to its velocity would depend on which probes we use to establish this lifetime?

What about Noninertial Observers?

The idea of DSR Minkowski limit was introduced [14] as possibility for one aspect of the quantum-gravity realm. Of course, from a full quantum-gravity perspective a description that is applicable only to the Minkowski limit represents only a small piece of the puzzle. Taking DSR as starting point for quantum gravity the next natural step would be to consider noninertial observers, a description of noninertial observers that is compatible with the DSR description of inertial observers. Or equivalently a description of curved (quantum) spacetimes, and a description of the dynamics of such spacetimes, which admits as Minkowski limit a DSR Minkowski limit. A key issue [98] for such studies comes from the observation that we might be required to attribute to the Planck scale a double role: a role in the gravitational coupling (because of the relation between G and the Planck scale) and a role in the structure of spacetime (and/or the structure of energy-momentum space). If this intuition turns out to be correct we might have to face significant challenges at the conceptual level. It is always very significant when two operatively well-defined (and apparently independent) physical quantities turn out to be identified (see, e.g., the Equivalence Principle for inertial and gravitational mass).

A Gravity Rainbow?

One of the many attempts to introduce mathematical structures suitable for the description of a DSR Minkowski limit is based on the introduction of an energy-dependent metric. This technique was named "gravity rainbow" by Magueijo and Smolin, who first proposed it as a tool for DSR research [54]. As for other formalisms being proposed for the description of a DSR Minkowski limit, also this proposal still lacks a fully developed interpretation and implementation. It is easy to see however how the language of an energy-dependent metric could be used for an effective description of modified energy-momentum relations in the Minkowski limit. For example, if the modified dispersion relation is of the type

$$\mathcal{C} = p^2 + f(E, E_p)E^2 \ , \tag{69}$$

where C is the deformed "mass Casimir" and E_p denotes the Planck energy scale, one could introduce an energy-dependent metric $g_{\mu\nu}(E, E_p)$ such that

$g_{00}(E, E_p) = f(E, E_p)\eta_{00}$ and $g_{ij} = \eta_{ij}$ and describe the same dispersion relation as

$$\mathcal{C} = p^\mu g_{\mu\nu}(E, E_p) p^\nu \ . \tag{70}$$

As mentioned above, the possibility of an observer-independent Planck-scale modification of the energy-momentum dispersion relation is one of the most studied possibilities as a "physics ingredient" for a DSR Minkowski limit, and therefore the gravity rainbow formulation might indeed be relevant for DSR research.

Actually, it is well known that in some DSR schemes, even some of those considered in the earliest DSR studies [14, 98], the abstraction of a classical spacetime (and of a standard metric) is not available in general. As stressed in [98], if an observer-independent nonlinear deformation of the dispersion relation is adopted and velocity is described by $v = dE/dp$, then one observer, "O", could see two particles with different masses m_A and m_B moving at the same speed and following the same trajectory (for O particles A and B are "near" at all times), while for a second observer O' the same two particles would have different velocities (so they could not possibly be "near" at all times). This then leads inevitably [98] to considering spacetime as an approximate concept, only valid within a certain class of observations and with a certain level of approximation. In the low-energy regime one could still introduce a spacetime and an associated metric. And one could still have [98] a spacetime and a metric (but a different metric, depending on the energy) in certain special high-energy processes (e.g. processes involving all particles with the same energy and mass). But in general (e.g. for processes involving several particles with large hierarchies of energies) one should do without [98] a conventional concept of spacetime (and of course without a conventional concept of metric).

So, at present, the "gravity rainbow" concept of an energy-dependent metric, while providing an intuitive characterization of modified dispersion relations and some associated effects, is still not naturally applicable to other aspects of a DSR Minkowski limit. Especially when several energy scales are involved, a simple-minded implementation of an energy dependence of the metric might lead to ambiguities.

One possibility that could perhaps be considered for a variant of the gravity-rainbow proposal consists [90] in replacing the energy-dependent metric with a corresponding statement of nonlinear relation between covariant fourmomentum and contravariant fourmomentum. After all (in an appropriate sense) the Minkowski limit does not really require us to make explicit reference to a metric: the ordinary $\eta_{\mu\nu}$ is only used to lower and raise indices, and in particular it is used to relate (linearly) the covariant fourmomentum and the contravariant fourmomentum. The energy dependence of the metric in the Minkowski limit could be a simple way to express a requirement of nonlinear relation between the covariant fourmomentum and the contravariant fourmomentum. One of the two (say, the covariant fourmomentum) could still transform according to ordinary Special Relativity, but then the relativistic properties of the other would codify departures from the special-relativistic predictions. This leads one to consider a

previously unexplored possibility for the construction of DSR Minkowski limits. Whereas usually in DSR research one assumes that the same nonlinear realization of the Lorentz symmetry group should be applied to all energy-momentum-space quantities, one should perhaps also contemplate the possibility that, say, the covariant fourmomentum still transforms linearly under Lorentz transformations, while the contravariant fourmomentum might indeed transform nonlinearly. In general one could notice that in the classical Minkowski limit various quantities, such as the covariant fourmomentum, the contravariant fourmomentum, and the frequency/wavenumber fourvector, all transform in the same linear way under Poincaré transformations, but in a Planck-scale-accurate description of the Minkowski limit some differences may arise, and, for example, the transformation rules of some of these quantities might still be linear, while some other of these quantities might transform nonlinearly.

4.11 On the Criteria for a DSR Minkowski Limit

For obvious "historic" reasons the quantum-gravity community is more accustomed to deal with mathematical frameworks rather than with physics concepts, and on some occasions the fact that for the physical concept of a DSR Minkowski limit we still do not have some fully satisfactory mathematical realizations has been source of some confusion. The tendency by some authors to identify the "physics project" of a DSR Minkowski limit with some specific formalism has also led to some inconsistency in the terminology. Additional confusion is generated by studies in which the authors quickly conclude that they are proposing a DSR Minkowski limit whenever "the Planck length takes the role of an absolute scale", without verifying that the "absolute scale" is such to require departures from some standard special-relativistic laws.

In light of this possibility of confusion it is perhaps useful to mention explicitly some possible roles for the Planck scale that would indeed require a DSR Minkowski limit, and some that would not. And let me start by observing that if in the Minkowski limit of a given quantum-gravity theory one had the Planck length setting an observer-independent minimum allowed value of wavelength, then of course one would be dealing with a DSR Minkowski limit. Under special-relativistic boosts wavelengths contract, and therefore in order to enforce an observer-independent minimum-wavelength law one should necessarily introduce departures from Special Relativity, and the observer independence of the postulated new law should allow to accommodate the departures from Special Relativity in such a way that the Galilei Relativity Principle would still hold. A role for the Planck length as observer-independent minimum wavelength would require a modification of Special Relativity just like one needs to modify Galilei Relativity in order to accommodate a maximum-speed law (speeds transform linearly under Galilei boosts).

Similarly a DSR Minkowski limit would necessarily arise in a quantum-gravity theory which in the Minkowski limit predicts the existence of some absolutely fundamental particles whose energy is constrained by an observer-independent bound ($E \leq 1/L_p$). But of course if in the Minkowski limit one

finds a bound on the mass (rest energy) of the particles then instead there is no a priori reason for expecting DSR structures. Mass is an invariant of Poincaré transformations, so an observer-independent bound on mass does not necessarily affect Poincaré symmetry. A useful example of the situation in which an absolute scale does not affect symmetries is provided by the Planck constant $\hbar$ in the quantum mechanics of angular momentum. Angular momentum transforms under space rotations, but $\hbar$ is most fundamentally a scale affecting the square modulus of angular momentum, $L_x^2 + L_y^2 + L_z^2$, which is an invariant under space rotations, and in fact the scale $\hbar$ can be introduced without affecting space-rotation symmetry [88].

For what concerns the fate of Poincaré symmetry in the Minkowski limit of quantum gravity it is therefore crucial to establish whether the Planck scale is introduced "*a la c*" (the speed bound introduced through c required a deformation of Galilei boosts) or is introduced "*a la* $\hbar$" (the properties of $L_x^2 + L_y^2 + L_z^2$ introduced through $\hbar$ do not affect in any way space-rotation symmetry). An early attempt to introduce a length scale (possibly the Planck length) in spacetime structure in such a way that it would not require any modification of Poincaré symmetry is the one of Snyder [86], who indeed postulated some spacetime noncommutativity and then provided a lengthy defense of the thesis that the system would still be Poincaré invariant. Some confusion may arise from the fact that in some recent papers (see, e.g., [111]) there has been some discussion of a "Snyder-type modification of special relativity", as this terminology misses the point that Snyder was trying to prove[23] just the opposite: an absolute length scale can be introduced without modifying Poincaré symmetry.

Going back to a list of concepts which imply or do not imply a DSR Minkowski limit, let me now focus on the minimum-uncertainty intuition that is common in the quantum-gravity literature. If in the Minkowski limit of quantum gravity there is an observer-independent bound on the measurability of lengths ($\delta L \geq L_p$) then necessarily this would have to be a DSR Minkowski limit. Instead there is no need to introduce departures from Poincaré symmetry if there is a bound on the measurability of proper lengths (the length of an object in its rest frame is of course a Poincaré-invariant quantity).

And in closing, since there is literature on the possibility of a maximum acceleration [112, 113], let me also stress that acceleration is a Poincaré invariant, and therefore a quantum-gravity theory predicting an observer-independent upper bound on acceleration will not require a DSR Minkowski limit.

While this is only a very limited list of examples it should suffice as a warning that it is not sufficient to argue that "the Planck length takes the role of an

[23] While Snyder should be credited for the idea of introducing the Planck scale in spacetime structure in such a way not to affect Pincaré symmetry, it is actually still unclear whether Snyder succeeded. Some of the tools more recently developed to analyze noncommutative geometries were not available to Snyder. Even now we only have a reliable description of rotation and boost transformations in the Snyder spacetime, which are indeed undeformed, whereas it is still unclear how to properly describe translations, which are often affected by severe ambiguities in noncommutative geometry (see, e.g., [7]).

absolute scale" in order to provide support for a DSR Minkowski limit. One must go through the (sometimes tedious, but extremely valuable) exercise of deriving the explicit form of the laws of transformation between observers, verifying that indeed the transformation laws are Planck-scale modified.

4.12 A DSR Challenge to the Wider Quantum-Gravity Community

The idea of DSR Minkowski limit also implicitly raises a challenge for those researchers in the quantum-gravity community who are seeking a "fundamental role for the Planck scale" without paying attention to the differences between the various types of fundamental scales that are possible in physics. For example, several papers adopt the hypothesis that the Planck length should set the minimum allowed value for wavelengths, but before the proposal in [14] of the concept of a DSR Minkowski limit this minimum-wavelength studies never explored the implications for special relativity. Similarly, there is a large literature on a vague hypothesis that the Planck length should set the absolute limit on the measurability of lengths, but the relevant studies often do not even provide an explicit statement concerning whether this absolute limit applies to the measurement of proper lengths or to the measurement of the length in any frame. The opposite attitude is equally dangerous for what concerns the amount of confusion produced in the literature: some authors, once they have established that in a chosen framework the Planck length has the role of "fundamental scale", quickly jump to the conclusion that they are dealing with a DSR Minkowski limit, whereas in order to draw such a conclusion one should first make sure that the Planck scale affects nontrivially the laws of transformation between inertial observers.

5 More on the Similarities with Beyond-Standard-Model Research

I have stressed already that there are some similarities between beyond-Special-Relativity research and beyond Standard-Model research. From the point of view of the development of a research methodology this represents an opportunity for beyond-Special-Relativity research, which should borrow heavily from the experience of the more mature field of beyond-Standard-Model research.

The key common point is that both research programmes look beyond some very successful theories, guided both by a general desire of further improving (extending) our confidence in the reliability of these theories and by an interest in exploring the implications of some arguments that appear to expose limitations in the applicability of these theories. For the Standard Model perhaps the best indication of incompleteness comes from the so-called "hierarchy problem", while for Special Relativity perhaps the best indication of incompleteness comes from arguments combining Quantum Mechanics and General Relativity that lead to the expectation that spacetime should not be describable as a smooth classical geometry at ultrashort distances. [In turn the expectation of a nonclassical

short-distance structure of spacetime leads to an expectation of departures from Special Relativity, since many properties of Special Relativity originate from the "smoothness" of the Minkowski spacetime.]

While some similarities in the structure of beyond-Special-Relativity research and beyond Standard-Model research have (probably fortuitously) emerged, including the mentioned amusing role that "loopholes" of the Coleman-Mandula theorem have played on both sides (with supersymmetric algebras and Hopf algebras), in general I feel that beyond-Special-Relativity research has not being profiting enough from the experience of beyond Standard-Model research. The situation is perhaps improving over the last couple of years, but traditionally (with a few noticeable exceptions) beyond-Special-Relativity research did not make use of the type of methodology that allows a phenomenology to make robust progress. Often the problems start already at the onset of the phenomenology, and may even involve unwise use of terminology. The most successful "test theories" beyond the standard model, such as the ones based on supersymmetry or technicolor, are well-defined proposals, consistently referred to throughout the particle-physics literature. In several instances beyond-Special-Relativity research has not relied on well-defined test theories, and it has happened that different articles ended up claiming different limits on the same parameters but were actually introducing those parameters within different test theories (making very different assumptions on how to introduce the relevant beyond-Special-Relativity parameters), and therefore one should have not even considered a comparison of the proposed experimental limits. I have presented elsewhere [43] a detailed case for the necessity to rely on some commonly-adopted test theories also on the beyond-Special-Relativity side.

Another example is the use of observations in astrophysics. Particle physicists use astrophysics observations very prudently, especially when (as it often happens) they are not yet well-understood/well-established. In many cases this ends up being taken perhaps too far, and it is assumed that one should only rely on laboratory data, disregarding completely astrophysics observations. For many aspects of beyond-Special-Relativity phenomenology there is no other option than the one of relying on astrophysics observations. If done cautiously, ensuring that the "experimental bounds" are derived in truly conservative manner, robust progress must be (perhaps slowly) achievable even using astrophysics observations, but too often beyond-Special-Relativity phenomenology has placed a lot of trust on astrophysics observations which were/are still not fully established and were/are still subject to many different interpretations.

Clearly a debate based on poorly specified test theories analyzed on the basis of some little-understood observations would not do much for the progress of beyond-Special-Relativity phenomenology. This phenomenology should and can do much better than this, as illustrated by several recent studies. Following the example of beyond-Standard-Model research is in this case particularly recommendable.

The careful use of terminology is also very important, especially in light of the sense of confusion with which, as mentioned, beyond-Special-Relativity research

is looked at from the outside. Terms such as "supersymmetry" and "technicolor" are used consistently throughout the particle-physics literature. Instead beyond-Special-Relativity proposals tend to dangerously change name rather frequently, and/or end up carrying multiple names. One example is the valuable research programme on the "Standard Model Extension": in several papers this name identifies a scheme whose primary ingredient (and a primary reason of appeal [114] for those working on it) is the power-counting renormalizability of the beyond-Special-Relativity corrections added to the Standard Model, but some other papers (at least some recent ones) are now reserving the name "Standard Model Extension" for a wider scheme, in which the power-counting-renormalizability requirement is removed, and denote as "Minimal Standard Model Extension" the one that insists on power-counting renormalizability. Somewhat similar is the situation for the use of the expression "varying-speed-of-light theories": at first this was used to characterize theories with a time-varying speed of light, and papers on the phenomenology of these varying-speed-of-light theories consistently focused on constraints for the time dependence of the speed of light, but recently the same name has also been used [17] to characterize the possibility of both a time- and an energy- dependence of the speed of light.

While the type of confusion induced by the use of names such as "Standard Model Extension" and "varying-speed-of-light theories" should probably be mostly armless, there are potentially more dangerous abuses of terminology in the beyond-Special-Relativity literature. In particular, a few recent papers (see, e.g., [111] and references therein) have renamed doubly special relativity as "deformed special relativity", and, besides the uneconomical and potentially confusing choice to carry two names for the same research programme, this is particularly inappropriate since there is an older research programme called "deformed special relativity" (see, e.g., [115] and references therein), which pursues completely different physics motivation and physics objectives.

Those readers who are using these notes as their first introduction to beyond-Special-Relativity research should be cautious in using keyword searches as a way to go deeper in the literature.

6 Another Century?

The specific beyond-Special-Relativity research line to which I devoted a disproportionately large part of these notes, the quantum-gravity-inspired doubly-special-relativity proposal, provides an example of how different strategies for the development of beyond-Special-Relativity research may fruitfully benefit form one another: in particular, all "general tests of Special Relativity" that preceded the proposal of doubly-special relativity were actually testing the hypothesis that both the first and second postulate would fail, the first most dramatically with the loss of equivalence of inertial observers. We are now challenged to contemplate tests of departures from Special Relativity that preserve the first (relativity-principle) postulate while allowing revisions the second postulate. This is therefore an idea that, even though it originated from considering

some aspects of the quantum-gravity problem, could be valuable for the whole physics community.

The assumption that the laws of Nature admit a "Minkowski limit" governed by Special Relativity can now claim more than a century of success. The fact that over these past few years on several types of modifications of Special Relativity the experimental limits have improved very rapidly, in spite of the fact that they failed to uncover any departures, can be seen as encouraging: the experimental side is reaching a certain maturity and if the pace of improvement continues to be so fast the discovery of departures from Special Relativity may really be behind the corner (or at least it might not require another century of patience). Also encouraging is the fact that the quality of the hints obtained from theoretical studies on quantum-gravity and cosmology has started to improve significantly. This may allow to focus and strengthen our efforts particularly in the directions suggested by the theory arguments.

In order to profit fully from these opportunities I feel that the methodology of phenomenological beyond-Special-Relativity work must be of the type that produces a robust intelligible link between theory and data and must establish "experimental limits" that are conservative and reliable enough to be used as "facts" by the rest of the physics community. To a large extent this is already being done, and, perhaps using as guidance the experience of beyond-Standard-Model research, one may be optimist that further improvements are forthcoming.

References

1. S. Majid and H. Ruegg, Phys. Lett. B334 (1994) 348.
2. J. Lukierski, H. Ruegg and W.J. Zakrzewski Ann. Phys. 243 (1995) 90.
3. G. Amelino-Camelia and S. Majid, hep-th/9907110, Int. J. Mod. Phys. A15 (2000) 4301.
4. J. Madore, S. Schraml, P. Schupp and J. Wess, hep-th/0001203, Eur. Phys. J. C16 (2000) 161.
5. A. Matusis, L. Susskind and N. Toumbas, hep-th/0002075, JHEP 0012 (2000) 002.
6. N.R. Douglas and N.A. Nekrasov, Rev. Mod. Phys. 73 (2001) 977.
7. G. Amelino-Camelia and M. Arzano, hep-th/0105120, Phys. Rev. D65 (2002) 084044; A. Agostini, G. Amelino-Camelia, F. D'Andrea, hep-th/0306013, Int. J. Mod. Phys. A19 (2004) 5187.
8. O. Bertolami and L. Guisado, hep-th/0306176, JHEP 0312 (2003) 013
9. R. Gambini and J. Pullin, Phys. Rev. D59 (1999) 124021.
10. J. Alfaro, H.A. Morales-Tecotl and L.F. Urrutia, Phys. Rev. Lett. 84 (2000) 2318.
11. T. Thiemann, gr-qc/0110034.
12. L. Smolin, hep-th/0209079.
13. G. Amelino-Camelia, L. Smolin and A. Starodubtsev, hep-th/0306134, Classical and Quantum Gravity 21 (2004) 3095.
14. G. Amelino-Camelia, gr-qc/0012051, Int. J. Mod. Phys. D11 (2002) 35; hep-th/0012238, Phys. Lett. B510 (2001) 255.
15. J.W. Moffat, Int. J. Mod. Phys. D2 (1993) 351.

16. A. Albrecht and J. Magueijo, Phys. Rev. D59 (1999) 043516; J.D. Barrow, Phys. Rev. D59 (1999) 043515.
17. J. Magueijo, astro-ph/0305457, Rept. Prog. Phys. 66 (2003) 2025.
18. J.K. Webb, V.V. Flambaum, C.W. Churchill, M.J. Drinkwater and J.D. Barrow, Phys. Rev. Lett. 82 (1999) 884.
19. M.T.Murphy et al, MNRAS 327 (2001) 1208.
20. T. Kifune, Astrophys. J. Lett. 518 (1999) L21.
21. R. Aloisio, P. Blasi, P.L. Ghia and A.F. Grillo, Phys. Rev. D62 (2000) 053010.
22. R.J. Protheroe and H. Meyer: Phys. Lett. B493 (2000) 1.
23. G. Amelino-Camelia and T. Piran, hep-ph/0006210, Phys. Lett. B497 (2001) 265; astro-ph/0008107, Phys. Rev. D64 (2001) 036005; G. Amelino-Camelia, gr-qc/0012049, Nature 408 (2000) 661.
24. J. Wess, lectures given at the "339th WE Herauser-Seminar: Special Relativity will it survive the next 100 years?" (Potsdam, 13–18 February 2005)
25. C. Lammerzahl, gr-qc/0402122, (Summary talk presented at 10th Marcel Grossmann Meeting on Recent Developments in Theoretical and Experimental General Relativity, Gravitation and Relativistic Field Theories, Rio de Janeiro, 20–26 July 2003)
26. G. Amelino-Camelia and F. Buccella, hep-ph/0001305, Mod. Phys. Lett. A15 (2000) 2119.
27. A. Apostolakis et al, hep-ex/9903005, Phys. Lett. B **452**, 425 (1999).
28. D. Colladay and V.A. Kostelecky, Phys. Rev. D55 (1997) 6760.
29. D. Colladay and V.A. Kostelecky, Phys. Rev. D58 (1998) 116002.
30. V.A. Kostelecky and C.D. Lane, Phys. Rev. D60 (1999) 116010.
31. A. Kostelecky and M. Mewes, Phys. Rev. D66 (2002) 056005.
32. P.L. Stanwix, M.E. Tobar, P. Wolf, M. Susli, C.R. Locke, E.N. Ivanov, J. Winterflood and F. van Kann, hep-ph/0506074, Phys. Rev. Lett. 95 (2005) 040404.
33. R. Bluhm, hep-ph/0506054.
34. W.-T. Ni, Phys. Rev. Lett. 38 (1977) 301.
35. M.P. Haugan and T.F. Kauffmann, Phys. Rev. D52 (1995) 3168.
36. G. Amelino-Camelia, J. Ellis, N.E. Mavromatos and D.V. Nanopoulos, hep-th/9605211, Int. J. Mod. Phys. A12 (1997) 607; G. Amelino-Camelia, J. Ellis, N.E. Mavromatos, D.V. Nanopoulos and S. Sarkar, astro-ph/9712103, Nature 393 (1998) 763.
37. R.C. Myers and M. Pospelov: hep-ph/0301124, Phys. Rev. Lett. **90**, 211601 (2003).
38. C. Lämmerzahl, Class. Quantum Grav. 14 (1998) 13.
39. C. Lämmerzahl, A. Macias, and H. Müller, gr-qc/0501048, Phys. Rev. D71 (2005) 025007.
40. G. Amelino-Camelia, C. Lammerzahl, A. Macias and H. Muller, gr-qc/0501053, AIP Conf. Proc. 758 (2005) 30.
41. G. Amelino-Camelia, G. Mandanici and K. Yoshida: hep-th/0209254, JHEP 0401 (2004) 037.
42. G. Amelino-Camelia, gr-qc/0402009.
43. G. Amelino-Camelia: gr-qc/0212002, New J. Phys. 6 (2004) 188.
44. H.P. Robertson, Rev. Mod. Phys. 21 (1949) 378.
45. R. Mansouri and R.U. Sexl, Gen. Rel. Grav. 8 (1977) 497.
46. P. Kosinski and P. Maslanka, hep-th/0211057, Phys. Rev. D68 (2003) 067702.
47. S. Mignemi, hep-th/0302065, Phys. Lett. A316 (2003) 173.
48. M. Daszkiewicz, K. Imilkowska and J. Kowalski-Glikman, hep-th/0304027.

49. J. Kowalski-Glikman, hep-th/0312140.
50. G. Amelino-Camelia, F. D'Andrea and G. Mandanici, hep-th/0211022, JCAP 0309 (2003) 006.
51. T. Jacobson, S. Liberati and D. Mattingly, hep-ph/0112207, Phys. Rev. D66 (2002) 081302.
52. J. Kowalski-Glikman, hep-th/0102098, Phys. Lett. A286 (2001) 391.
53. R. Bruno, G. Amelino-Camelia and J. Kowalski-Glikman, hep-th/0107039, Phys. Lett. B522 (2001) 133.
54. J. Magueijo and L. Smolin, hep-th/0112090, Phys. Rev. Lett. 88 (2002) 190403; gr-qc/0207085, Phys. Rev. D67 (2003) 044017; gr-qc/0305055, Class. Quant. Grav. 21 (2004) 1725.
55. S. Judes and M. Visser, gr-qc/0205067, Phys. Rev. D68 (2003) 045001;
56. G. Amelino-Camelia, D. Benedetti, F. D'Andrea, hep-th/0201245, Class. Quant. Grav. 20 (2003) 5353; J. Kowalski-Glikman and S. Nowak, hep-th/0204245, Int. J. Mod. Phys. D12 (2003) 299; D. Kimberly, J. Magueijo and J. Medeiros, gr-qc/0303067.
57. S. Alexander and J. Magueijo, hep-th/0104093.
58. Martin Bojowald, Phys. Rev. Lett. 86 (2001) 5227.
59. A. Ashtekar, M. Bojowald and J. Lewandowski, Adv. Theor. Math. Phys. 7 (2003) 233.
60. Martin Bojowald, R. Maartens and P. Singh, Phys. Rev. D70 (2004) 083517.
61. C. Barcelo, S. Liberati and M. Visser, gr-qc/0505065.
62. L.J. Garay, J.R. Anglin, J.I. Cirac and P. Zoller Phys. Rev. Lett. 85 (2000) 4643.
63. M. Visser, C. Barcelo and S. Liberati, Gen. Rel. Grav. 34 (2002) 1719.
64. C. Rovelli, gr-qc/9710008, Living Rev. Rel. 1 (1998) 1.
65. L. Smolin, hep-th/0303185; L. Smolin, "Three Roads to Quantum Gravity" (Weidenfeld and Nicolson, London, 2000).
66. S. Doplicher, K. Fredenhagen and J.E. Roberts, Phys. Lett. B331 (1994) 39.
67. D.V. Ahluwalia, Phys. Lett. B339 (1994) 301.
68. Y.J. Ng and H. Van Dam, Mod. Phys. Lett. A9 (1994) 335.
69. G. Amelino-Camelia, gr-qc/9603014, Mod. Phys. Lett. A9 (1994) 3415; gr-qc/9603013; Mod. Phys. Lett. A11 (1996) 1411.
70. L.J. Garay, Int. J. Mod. Phys. A10 (1995) 145.
71. G.E. Volovik, gr-qc/0301043, Found. Phys. 33 (2003) 349.
72. G. Chapline, E. Hohlfeld, R.B. Laughlin and D.I. Santiago, Phil. Mag. 81 (2001) 235; gr-qc/0012094; R.B. Laughlin, gr-qc/0302028, Int. J. Mod. Phys. A18 (2003) 831.
73. J. Madore, S. Schraml, P. Schupp and J. Wess, hep-th/0001203, Eur. Phys. J. C16 (2000) 161.
74. P. Kosinski, J. Lukierski and P. Maslanka, Czech. J. Phys. 50 (2000) 1283.
75. A. Anisimov, T. Banks, M. Dine and M. Graesser hep-ph/0106356.
76. G. Amelino-Camelia, L. Doplicher, S. Nam and Y.-S. Seo, hep-th/0109191, Phys. Rev. D67 (2003) 085008.
77. M. Dimitrijevic, L. Jonke, L. Moller, E. Tsouchnika , J. Wess and M. Wohlgenannt, hep-th/0307149, Eur. Phys. J. C31 (2003) 129.
78. M. Chaichian, P.P. Kulish, K. Nishijima and A. Tureanu, Phys. Lett. B604 (2004) 98.
79. J. Wess, hep-th/0408080.
80. C. Gonera, P. Kosinski, P. Maslanka and S. Giller, Phys. Lett. B622 (2005) 192.

81. G. Veneziano, Europhys. Lett. 2 (1986) 199; D.J. Gross and P.F. Mende, Nucl. Phys. B303 (1988) 407; D. Amati, M. Ciafaloni, G. Veneziano, Phys. Lett. B216 (1989) 41; K. Konishi, G. Paffuti, P. Provero, Phys. Lett. B234 (1990) 276; T. Yoneya, Mod. Phys. Lett. A4 (1989) 1587.
82. D. Kabat and P. Pouliot, Phys. Rev. Lett. 77 (1996) 1004.
83. M.R. Douglas, D. Kabat, P. Pouliot, S.H. Shenker, Nucl. Phys. B485 (1997) 85.
84. E. Witten, Phys. Today 49 (1996) 24.
85. C. Rovelli and L. Smolin, Nucl. Phys. B442 (1995) 593.
86. H.S. Snyder, Phys. Rev. 71 (1947) 38.
87. C. Rovelli and S. Speziale, gr-qc/0205108.
88. G. Amelino-Camelia, gr-qc/0205125; gr-qc/0309054.
89. L. Freidel, J. Kowalski-Glikman and L. Smolin, hep-th/0307085.
90. G. Amelino-Camelia, gr-qc/0506117.
91. N.E. Mavromatos, hep-ph/0309221.
92. A. Agostini, PhD thesis (Univ Naples, Italy), hep-th/0312305.
93. A. Agostini, G. Amelino-Camelia and M. Arzano, gr-qc/0207003, Class. Quant. Grav. 21 (2004) 2179.
94. D.V. Ahluwalia, gr-qc/0207004.
95. G. Amelino-Camelia: gr-qc/0107086, Phys. Lett. B **528**, 181 (2002).
96. T.J. Konopka and S.A. Major: New J. Phys. **4**, 57 (2002).
97. O. Bertolami: hep-ph/0301191.
98. G. Amelino-Camelia, gr-qc/0210063, Int. J. Mod. Phys. D11 (2002) 1643.
99. G. Amelino-Camelia, astro-ph/0209232, Int. J. Mod. Phys. D12 (2003) 1211.
100. J. Lukierski, H. Ruegg, A. Nowicki and V.N. Tolstoi, Phys. Lett. B264 (1991) 331; J. Lukierski, A. Nowicki and H. Ruegg, Phys. Lett. B293 (1992) 344.
101. A. Nowicki, E. Sorace and M. Tarlini, hep-th/9212065, Phys. Lett. B302 (1993) 419.
102. P.N. Bibikov, q-alg/9710019, J. Phys. A31 (1998) 6437.
103. P. Kosinski, P. Maslanka, J. Lujierski and A. Sitarz, Czech. Journ. Phys. 48 (1998) 11.
104. P. Kosinski, P. Maslanka and J. Lujierski, hep-th/0103127.
105. V.G. Drinfeld, "Quantum Groups", Proceedings of the International Congress of Math. Berkeley, USA (1986), p. 793; M. Jimbo, Lett. Math. Phys. 10 (1985) 63; M. Jimbo, Lett. Math. Phys. 11 (1986) 247.
106. J. Lukierski, H. Ruegg and W. Ruhl, Phys. Lett. B313 (1993) 357.
107. I.A. Batalin, J. Math. Phys. 22 (1981) 1837.
108. G. Amelino-Camelia, J. Kowalski-Glikman, G. Mandanici and A. Procaccini: gr-qc/0312124.
109. J. Lukierski and A. Nowicki, hep-th/0203065.
110. G. Amelino-Camelia, in preparation.
111. F. Girelli and E.R. Livine, gr-qc/0412004.
112. E.R. Caianiello, Riv. Nuovo Cim. 15 (1992) 1; E.R. Caianiello, M. Gasperini and G. Scarpetta, Nuovo Cim. 105B (1990) 259.
113. F.P. Schuller, M.N.R. Wohlfarth and T.W. Grimm, hep-th/0211264, Class. Quantum Grav. 20 (2003) 3947; F.P. Schuller and Hendryk Pfeiffer, hep-th/0307247.
114. R.L. Walsworth, D. Bear, M. Humphrey, E.M. Mattison, D.F. Phillips, R.E. Stoner and R.F.C. Vessot, AIP Conf. Proc. 539 (2000) 119; V.W. Hughes, M. Grosse Perdekamp, D. Kawall, W. Liu, K. Jungmann and G. zu Putlitz, Phys. Rev. Lett. 87 (2001) 111804.
115. F. Cardone and R. Mignani, Grav. and Cosm. 4 (1998) 311; F. Cardone, A. Marrani and R. Mignani, Found. Phys. Lett. 16 (2003) 163.

Doubly Special Relativity as a Limit of Gravity

K. Imiłkowska and J. Kowalski-Glikman

Institute of Theoretical Physics, University of Wrocław, Pl. Maxa Borna 9, 50-204 Wrocław, Poland
kaim@ift.uni.wroc.pl, jurekk@ift.uni.wroc.pl

1 Introduction

The anniversary of a great idea is usually a good occasion for critical reassessment of its meaning, influence, and future. In theoretical physics, where methodology, instead of hermeneutics, is based on Popperian conjectures and refutations scheme this last issue – the future – is, of course, the most important. Thus, in the course of the celebrations of the 100 anniversary of the Theory of Relativity, we are mostly interested in asking the questions: Is Special Relativity still to be regarded as the correct theory describing relativistic phenomena (particles and fields kinematics and dynamics) in flat space-time? Will it survive the next 100 years, and if not, which theory is going to replace it?

One quite often hears the opinion that there is, in fact, no such theory as Special Relativity. What we have to do with is just a very particular, flat space-time limit of General Relativity. And given the fact that few of us doubt that the ultimate theory of gravity should be Quantum Gravity (in the form of Loop Quantum Gravity, or String Theory, or perhaps – and most likely – in the disguise we do not really know yet) the question to be posed is: what is the flat space, semiclassical limit of Quantum Gravity?

For a long time it was taken as obvious that such a limit should be just the ordinary Special Relativity. Recent developments, however, put some doubts on this naive conclusion. First, assuming that one or another form of (super)strings theory is indeed the correct theory of Quantum Gravity one may contemplate the idea that in the (super)string vacuum, corresponding to our universe, Lorentz invariance is spontaneously broken. This would certainly lead to some, possibly observable, modifications of Special Relativity [1]. Violation of Lorentz symmetry is also possible in models based on Loop Quantum Gravity [2]. The issue of Lorentz Invariance Violation and its possible observational consequences has been recently reviewed in [3] and [4].

The core of all the proposals of spontaneous Lorentz symmetry breaking is introduction of some sort of aether. While phenomenological models of this form are certainly interesting, making possible to device precise experimental tests,

K. Imiłkowska and J. Kowalski-Glikman: *Doubly Special Relativity as a Limit of Gravity*, Lect. Notes Phys. **702**, 279–298 (2006)
DOI 10.1007/3-540-34523-X_10
© Springer-Verlag Berlin Heidelberg 2006

they are much less appealing theoretically, since they are based on rejection of the most cherished principle of physics, the relativity principle.[1] In our view there is no a priori reason, neither theoretical, nor experimental, to contemplate violation of relativity principle. This does not mean that modifications of Special Relativity are not possible, and we will argue below that there are good reasons to believe that if one regards Special Relativity (understood as a kinematical theory of particles and fields on flat semiclassical space-time) as a limit of (quantum) gravity, it is inevitable that some, Planck scale corrections, must be present. In a sense, the presence of this scale reflects the "memory" of flat space-time about its (quantum) gravitational origin.

2 Postulates of Doubly Special Relativity

Let us start with the plausible assumption that the flat space-time kinematics of point particles originates in some theory of quantum gravity. Assume further that in taking the flat space-time, semiclassical limit there is no breaking of spacetime symmetries, and in particular that the relativity principle holds (in other words, the diffeomorphism invariance of the theory is not being broken in the process.) After taking this limit the space-time will be, at least locally, the standard Minkowski space-time, however it may well be that the resulting theory still possesses some information about its origin, in the form of the observer independent mass scale, κ, of order of Planck scale. This scale will be still present as a parameter in the transformation laws, relating different inertial observers.

If such a scale indeed is present, it is natural to expect that deviations from the standard Special Relativistic kinematics arise in the processes characterized by the energy scale close to κ, and that these deviations should be rather generic [5]. For example, it may happen that the standard dispersion relation for particles acquires additional terms, to wit

$$E^2 = p^2 + m^2 + \alpha \frac{E^3}{\kappa} + \dots \tag{1}$$

with α being a dimensionless parameter of order 1.

Naively, modified dispersion relation would immediately imply the existence of preferred frame, since if it holds in one frame it does not hold in any other,

[1] One should note that the presence of aether does not necessarily mean breaking of Lorentz symmetry. It is well possible that the relations between inertial observers still form Lorentz (or Poincaré) group, however, group elements would, in the presence of the aether, depend both on relative velocity of observers *and* velocity with respect to the aether. Physically, the breaking of Lorentz symmetry means that some physical processes (particle reactions, for example) are possible for some range of velocities relative to the aether and impossible for other velocities. In DSR proposal, which we will describe below, the central postulate is that relativity principle still holds, and thus if some process is observed by one observer it is observed by all other. There is also some specific relation between descriptions of the same process by two observers, which depends *only* on their relative velocity.

related to the original one by the standard Lorentz transformation. However we can demand that along with deforming dispersion relation we deform Lorentz transformations, so that (1) (and their generalizations to be introduced shortly) remain invariant under action of six parameter group of transformations, with generators satisfying Lorentz $SO(3,1)$ algebra.

Based on this intuition, Giovanni Amelino-Camelia [6, 7] formulated a set of postulates that should be satisfied by new theory, replacing Special Relativity, in the regime of ultra-high energies, and he dubbed this theory Doubly Special Relativity.

Doubly Special Relativity is based on the following postulates:

1. *The Relativity Principle holds.* This means that if two observers describe the same phenomenon, possible differences in their descriptions can only depend on their relative motion (in the case of inertial observers – their relative velocity). In particular there is no notion of absolute rest, and absolute motion. It follows also that if some process is observed by one observers (for example particle (1) collides with particle (2) producing particle (3)), all observers agree that this process takes place.
2. *There exist two observer-independent scales*: the velocity scale c, identified with velocity of light, and, the mass scale κ, identified with Planck mass ($\kappa \sim 10^{19}$ GeV).

It is worth recalling at this point what is the difference between observer independent scales and other dimensionful quantities one encounters in physical theories. To be more specific, let us consider the question what is the difference, in the framework of the standard Special Relativity, between a coupling constant, like electric charge e or Planck's constant $\hbar$, and the velocity of light c. In the first case, of dimensionful coupling constants, all observers measure their values with the help of identical, low energy experiments, performed in their rest frame. Then the relativity principle[2] guarantees that all the observers will obtain the same numerical values of the constants. These values can be then used in other experiments. In the case of the speed of light the situation is different however. Since this speed is an observer independent scale, we demand also that if two inertial observers measure velocity of *the same* photon, they will obtain the same result. This is clearly incompatible with Galilean Relativity, but, as we know, has been successfully built into the principles of his Special Relativity. Analogously, in the construction of Doubly Special Relativity we postulate the presence of yet another observer-independent scale, this time of dimension of mass, whose numerical value is presumed to be of order of Planck mass.[3]

[2] In the "passive" form: Identical experiments performed by inertial observers give the same results.

[3] Note, however, that contrary to Special Relativity, in which we know physical objects moving with the velocity of light (massless particles), in the case of DSR we do not know (yet?) physical objects exemplifying the scale κ.

It has soon been realized [8,9] (for recent review see [10]) that these postulates are satisfied in the framework of theories in which Poincaré algebra is replaced by deformed κ-Poincaré algebra [11–14].[4]

One possible realization is provided by the, so-called, bicrossproduct (or Majid–Ruegg) basis [13], in which the commutators between rotation M_i, boost N_i, and momentum $P_\mu = (P_0, P_i)$ generators are the following

$$[M_i, M_j] = \epsilon_{ijk} M_k, \quad [M_i, N_j] = \epsilon_{ijk} N_k,$$
$$[N_i, N_j] = -\epsilon_{ijk} M_k \ . \tag{2}$$

and

$$[M_i, P_j] = \epsilon_{ijk} P_k, \quad [M_i, P_0] = 0$$
$$[N_i, P_j] = \delta_{ij} \left(\frac{1}{2} \left(1 - e^{-2P_0/\kappa} \right) + \frac{\mathbf{P}^2}{2\kappa} \right) - \frac{1}{\kappa} P_i P_j \ ,$$
$$[N_i, P_0] = P_i \ .$$

As one can easily check, the Casimir of the κ-Poincaré algebra (2), (3) reads

$$\mathcal{C} = \kappa^2 \cosh \frac{P_0}{\kappa} - \frac{\mathbf{P}^2}{2} e^{P_0/\kappa} - M^2. \tag{3}$$

It is easy to check also that the expansion of (3) to the leading order in $1/\kappa$ yields (1).

One should note at this point that the bicrossproduct algebra above is not the only possible realization of DSR. For example, in [15, 16] Magueijo and Smolin proposed and carefully analyzed another DSR proposal, called sometimes DSR2. In DSR2 the Lorentz algebra is still not deformed and there are no deformations in the brackets of rotations and momenta. The boosts– momenta generators have now the form

$$[N_i, p_j] = i \left(\delta_{ij} p_0 - \frac{1}{\kappa} p_i p_j \right), \tag{4}$$

and

$$[N_i, p_0] = i \left(1 - \frac{p_0}{\kappa} \right) p_i. \tag{5}$$

It is easy to check that the Casimir for this algebra has the form

$$M^2 = \frac{p_0^2 - \mathbf{p}^2}{\left(1 - p_0/\kappa\right)^2}. \tag{6}$$

In order to describe kinematics of a particle we must extend the above algebra to the algebra of phase space of the particle. We will not derive here all the results, and the reader could find the derivation with references to the original

[4] It should be stressed however that DSR is not just the κ-Poincaré algebra – not only in the sense analogous to the well known fact Special Relativity is not just the Poincaré algebra – there might be DSR proposals in which this algebra does not play any role.

literature in the review paper [10]. Instead we will just state the main results which will be important below, when we compare DSR with a theory resulting as a flat limit of gravity coupled to point particles.

- Both the bicrossproduct (2)–(3) and Magueijo–Smolin algebras (4)–(6) can be understood as examples of larger class of algebras constructed as follows. In the standard Special Relativity four-momentum can be thought of as a point of four dimensional flat manifold of Lorentz Signature – the flat momentum manifold. Notice that in this case positions, being "translations of momenta" are points of another flat Minkowski space. In transition to DSR, assume instead that the space of momenta is a maximally symmetric manifold of (constant) curvature, $1/\kappa^2$. In the limit when κ goes to infinity, the curvature goes to zero and we return to Special Relativity, as we should. In given coordinates on this constant curvature momentum space, each point will correspond to some four-momentum. As it is well known, the group of symmetries in this case is the 10-parameter (in 4 dimensions) de Sitter group $\mathsf{SO}(4,1)$. This group possesses a six-parameter subgroup, $\mathsf{SO}(3,1)$, isomorphic with Lorentz group, and one can easily compute what will be infinitesimal action of the group elements on points of the manifold. It turns out that, for example, the bicrossproduct basis corresponds to the standard system of coordinates, used in cosmology. More details can be found in [10] and [17].
- What about the remaining four parameters of de Sitter group $\mathsf{SO}(4,1)$? It is well known from differential geometry, that while the generators of $\mathsf{SO}(3,1)$ act as "rotations" the remaining ones play the role of "translations". This means that it is natural to identify them with positions. It turns out to be convenient to arrange the remaining four generators x^μ so as to form the Iwasawa decomposition of the $\mathsf{so}(4,1)$ algebra; explicitly their commutators could be brought to the following form

$$\left[x^i, x^j\right] = 0, \quad \left[x^0, x^i\right] = \frac{1}{\kappa} x^i \tag{7}$$

  The noncommutative space-time satisfying (7) is called κ-Minkowski space-time.
- It should be noted that there is a natural Hopf algebra structure associated with an algebra of symmetries of de Sitter space. This algebra turns out to be exactly the quantum κ-Poincaré algebra of [11–14]. For more details see [18].
- It should be also noted that one can in principle construct analogous structure starting from the momentum space of the particle being anti–de Sitter space [19]. Explicit models in four dimensions are not known in this case, however they play a role in 2+1 gravity coupled to a particle, as we will discuss in details below.

Given characterization of properties of single particle DSR models let us now turn to the question, what is DSR coming from.

3 Constrained BF Action for Gravity

It is usually considered rather obvious that Special Relativity, regarded as a theory of particle kinematics should emerge somehow from General Relativity coupled to point particles in a limit, in which gravitational interactions are "switched off". It turns out that it is surprisingly difficult to prove this claim in the framework of the standard Einstein formulation of GR. First of all "switching off" gravity would presumably mean going to zero with gravitational constant, but this limit is known to be pathological in GR. Moreover it is well known that coupling of GR to point particle is at least problematic, if not impossible.

Our starting point must be therefore some another (but equivalent) form of the gravity action. The convenient form has been derived recently by Freidel and Starodubtsev [20]. The starting point of that paper is the observation theory of gravity can be defined by the action containing two parts: the "vacuum" one, being a topological field theory with an appropriate gauge group, and the constraints, leading to the emergence of the dynamical degrees of freedom of gravity. Both parts are manifestly diffeomorphism invariant, which opens new perspectives in construction of diffeomorphism-invariant perturbation theory for quantum gravity. From our perspective, however, the most important aspect of this theory would be that it makes it possible both to define a limit, in which local degrees of freedom of gravity are switched off and the point particle coupling. One can say that in this formulation theory of gravity has the well defined "DSR limit."

The construction of the Freidel–Starodubtsev theory borrows from earlier works [21–23] and is based on the $\mathsf{SO}(4,1)$ gauge theory. The basic dynamical variables are[5] $\mathsf{so}(4,1)$ connection one form $\mathbf{A}^{IJ}$, and the $\mathsf{so}(4,1)$-valued two-form $\mathbf{B}^{IJ}$. The starting point is the action principle [20]

$$S = \int \mathbf{B}^{IJ} \wedge \mathbf{F}_{IJ} \tag{8}$$

where $\mathbf{F}_{IJ}$ is the curvature of connection $\mathbf{A}^{IJ}$. The equations of motion following from this action

$$\begin{aligned} \mathbf{F}_{IJ} &= 0 \\ d_{\mathbf{A}}\mathbf{B}_{IJ} &= 0 \end{aligned} \tag{9}$$

where $d_{\mathbf{A}}$ is the covariant derivative of connection $\mathbf{A}$, tell that the connection is flat, while the $\mathbf{B}^{IJ}$ field is covariantly constant. The solutions of these equations on locally connected region $\mathcal{U}$ is of the form

$$\mathbf{A} = \mathsf{g}^{-1}\, d\,\mathsf{g} \quad \mathbf{B} = \mathsf{g}^{-1}\, d\mathsf{f}\,\mathsf{g}, \quad \mathsf{g} \in \mathsf{SO}(4,1), \;\; \mathsf{f} \in \mathsf{so}(4,1) \tag{10}$$

The theory is therefore almost trivial, without any local degrees of freedom.

[5] Below we use **BOLD** typeface to denote forms (space-time indices suppressed) and SANS SERIF typeface to denote Lie algebra valued fields (group indices suppressed).

In order to get General Relativity we must break local symmetry of the theory from $\mathsf{SO}(4,1)$ down to the Lorentz group $\mathsf{SO}(3,1)$. To this end we denote by 5 the preferred direction in the algebra space, and add to the action the term which explicitly breaks the $\mathsf{SO}(4,1)$ gauge symmetry, to wit

$$S = \int \mathbf{B}^{IJ} \wedge \mathbf{F}_{IJ} - \frac{\alpha}{2} \mathbf{B}^{IJ} \wedge \mathbf{B}^{KL} \epsilon_{IJKL5} \tag{11}$$

Let us now decompose the algebra index $I = (i, 5)$, $i = 0, \ldots, 3$ with $\epsilon^{ijkl} = \epsilon^{IJKL5}$ being an invariant $\mathsf{SO}(3,1)$ tensor. Note that now the first equation in (9) is replaced by

$$\mathbf{F}_{IJ} = \alpha \, \mathbf{B}^{KL} \epsilon_{IJKL5} \tag{12}$$

and is manifestly *not* $\mathsf{SO}(4,1)$ covariant.

The **B**-field enters the action S only algebraically, so we can substitute the solution (12) back to the action (11) to get

$$S = \frac{1}{4\alpha} \int \mathbf{F}^{ij} \wedge \mathbf{F}^{kl} \epsilon_{ijkl}. \tag{13}$$

It is convenient at this point to decompose the curvature as follows

$$\begin{aligned} \mathbf{F}^{ij}(\mathbf{A}) &= \mathbf{R}^{ij}(\boldsymbol{\omega}) - \frac{1}{l^2} \mathbf{e}^i \wedge \mathbf{e}^j \\ \mathbf{F}^{i5}(\mathbf{A}) &= \frac{1}{l} d_{\boldsymbol{\omega}} \mathbf{e}^i \end{aligned} \tag{14}$$

where $\boldsymbol{\omega}^{ij} = \mathbf{A}^{ij}$ is the 4-dimensional connection one-form and $\mathbf{e}^i = e_\mu{}^i \, dx^\mu$ is a frame field which corresponding to the metric $g_{\mu\nu} = e^i_\mu \, e_{i\nu}$. $\mathbf{R}^{ij}$ is the $\mathsf{so}(3,1)$ curvature of connection $\boldsymbol{\omega}$, $\mathbf{R}^{ij}(\boldsymbol{\omega}) = d\boldsymbol{\omega}^{ij} + \boldsymbol{\omega}^i_k \wedge \boldsymbol{\omega}^{kj}$. Notice that for dimensional reasons we had to introduce the scale l, of dimension of length. Using the equations for the curvature (14), we can rewrite the action in terms of $\mathsf{so}(3,1)$ curvature:

$$\begin{aligned} S &= \frac{1}{4\alpha} \int (\mathbf{R}^{ij}(\boldsymbol{\omega}) - \frac{1}{l^2} \mathbf{e}^i \wedge \mathbf{e}^i) \wedge (\mathbf{R}^{kl}(\boldsymbol{\omega}) - \frac{1}{l^2} \mathbf{e}^k \wedge \mathbf{e}^l) \epsilon_{ijkl} \\ &= -\frac{1}{2G} \int (\mathbf{R}^{ij}(\boldsymbol{\omega}) \wedge \mathbf{e}^k \wedge \mathbf{e}^l - \frac{\Lambda}{6} \, \mathbf{e}^i \wedge \mathbf{e}^k \wedge \mathbf{e}^l) \epsilon_{ijkl} \\ &\quad + \frac{1}{4\alpha} \int \mathbf{R}^{ij}(\boldsymbol{\omega}) \wedge \mathbf{R}^{kl}(\boldsymbol{\omega}) \epsilon_{ijkl} \end{aligned} \tag{15}$$

What we get is nothing but the Palatini action of General Relativity with cosmological constant plus additional term whose variation vanishes identically due to Bianchi identity. Note that the Newton's constant G equals αl^2, while the cosmological constant $\Lambda = 3/l^2$. Thus the coupling constant $\alpha = G\Lambda/3$ is dimensionless and extremely small, which makes it a perfect candidate for a parameter of (both classical and quantum) perturbative expansion. As stressed by Freidel and Starodubtsev [20], the constrained BF theory is therefore very promising as a starting

point for construction of perturbative quantum gravity, where diffeomorphism invariance is manifestly preserved at all steps of perturbative expansion.

To the initial, topological action (8) we can still add the $\mathsf{SO}(4,1)$ cosmological term of the form

$$-\frac{\beta}{2}\int \mathbf{B}^{IJ}\wedge\mathbf{B}_{IJ}\ . \tag{16}$$

This addition changes the equations of motion:

$$\begin{aligned}\mathbf{F}_{IJ}-\beta\,\mathbf{B}_{IJ}\ &=0\\ d_{\mathbf{A}}\mathbf{B}_{IJ}&=0\end{aligned} \tag{17}$$

(the second equation follows in fact from the first and Bianchi identity.)

It can be shown that the action (16) is still topological, i.e., without local degrees of freedom. As before we can add to this action the α constraint, in order to obtain the action of General Relativity with the additional term $\frac{2}{\gamma}\,\mathbf{R}^{ij}(\boldsymbol{\omega})\wedge\mathbf{e}_i\wedge\mathbf{e}_j$ and more more topological terms. The "bare action" parameters l,α,β are related to the physical ones G, Λ, and γ (Immirzi parameter) as follows $\Lambda=3/l^2$, $\gamma=\beta/\alpha$, $G=\frac{\alpha^2-\beta^2}{\alpha}\,l$ (for more details and discussion of possible physical relevance of γ parameter see [20], and also recent paper [24].)

The convenient basis of $\mathsf{so}(4,1)$ algebra is provided by Dirac matrices $\gamma^{ij}=\frac{1}{2}[\gamma^i,\gamma^j]$ and $\gamma^i\gamma^5$. Using the $\mathsf{so}(4,1)$ algebra valued fields A_μ, $\mathsf{B}_{\mu\nu}$ the constrained BF action for gravity can be rewritten in the following form

$$\begin{aligned}S=&\int d^4x\epsilon^{\mu\nu\rho\sigma}Tr(\mathsf{B}_{\mu\nu}\mathsf{F}_{\rho\sigma}(\mathsf{A}))\\ &-\frac{\beta}{2}\int d^4x\epsilon^{\mu\nu\rho\sigma}Tr(\mathsf{B}_{\mu\nu}\mathsf{B}_{\rho\sigma})\\ &-\frac{\alpha}{2}\int d^4x\epsilon^{\mu\nu\rho\sigma}Tr(\mathsf{B}_{\mu\nu}\mathsf{B}_{\rho\sigma}\gamma^5)\end{aligned} \tag{18}$$

It is quite easy to couple point particles to the constrained BF action. Indeed since $\mathbf{A}^{IJ}$ is a one form, it couples naturally to one-dimensional objects – the particles world-lines.

The general procedure of coupling particles carrying non-abelian charges to Yang-Mills potential has been developed by Balachandran, Marmo, Skagerstam, and Stern (see [25] and references therein.) In the case at hands[6] the gauge group is $\mathsf{SO}(4,1)$. This group acts by conjugation on its algebra, and the orbits can be labelled by two numbers, corresponding to values of two Casimirs, representing mass and spin of the particle, as follows

$$\mathsf{K}=\frac{1}{2}\,m\gamma_1\gamma^5+\frac{1}{4}\,s\,\gamma_2\gamma_3 \tag{19}$$

As the second ingredient we take connection, gauge-transformed by an arbitrary element h of the Lorentz subgroup $\mathsf{SO}(3,1)$ of $\mathsf{SO}(4,1)$

[6] The results presented below have been obtained in collaboration with L. Freidel and A. Starodubtsev.

$$\mathsf{A}_\mu{}^{\mathsf{h}} = \mathsf{h}^{-1}\mathsf{A}_\mu\mathsf{h} + \mathsf{h}^{-1}\partial_\mu\mathsf{h}, \quad \mathsf{h} = \exp\left(\frac{1}{4}\,\alpha^{ab}\gamma_{ab}\right) \tag{20}$$

where, as before $\mathsf{so}(4,1)$ connection A decomposes into $\mathsf{so}(3,1)$ connection ω and tetrad e

$$\mathsf{A}_\mu = \left(\frac{1}{2}\,e_\mu{}^a\,\gamma_a\gamma^5 + \frac{1}{4}\omega_\mu{}^{ab}\,\gamma_{ab}\right) \tag{21}$$

Then the action of the particle with mass m and spin s coupled to constrained BF gravity is defined to be

$$L(z,h) = Tr\,\left(\mathsf{K}\mathsf{A}_\tau{}^{\mathsf{h}}(\tau)\right) \quad S = \int d\tau\, L\,, \tag{22}$$

where $\mathsf{A}_\tau{}^{\mathsf{h}} \equiv \mathsf{A}_\mu{}^{\mathsf{h}}(z(\tau))\dot{z}^\mu(\tau)$ is the value of gauge transformed connection (20) on the particle world-line. We see therefore that the dynamics of the particle is described with the help of the charge K it carries, and the Lorentz transformation h relating the particle rest frame and the frame of (asymptotic) observer. It can be shown that variation of the action (22) leads to the correct Mathiasson-Papapetrou equations describing the dynamic of spinning particle in the presence of torsion. When the torsion is zero we recover the usual Mathiasson-Papapetrou equation, which in the case of vanishing spin reduces to the usual geodesic equation.

Having defined the coupling of the particle to gravitational field we can address the question as to what would be the effective behavior of the particle in the limiting case, when the local degrees of freedom of gravitational field are being switched off. To answer this question, one should proceed as follows: take the action being the sum of (18) and (22), solve the resulting equations of motion, and then take the limit $\alpha \to 0$. To see which outcomes of this procedure are possible, note that that although in this limit the gravitational field will become flat in the bulk space-time, there might be some nontrivial leftover at the worldline of the particle. This contribution of the gravitational field may lead to deformation of the (otherwise free) particle action (22), leading to DSR like behavior.

Unfortunately, due to the technical difficulties, the programme described above has not been realized in practice yet. What can be done, however, is to turn to a simpler model of gravity coupled to a particle, in 2+1 dimensions. As we will see in the next section the structure of this model is very similar to the four-dimensional case with the parameters α, β equal zero, i.e., in the limit we are mostly interested in. Moreover, the three-dimensional case is not purely of academic interest, as the following argument, borrowed from [26] and [10], clearly shows.

The main idea is to construct an experimental situation that forces a dimensional reduction from the four dimensional to the $2+1$ dimensional theory. It is interesting that this can be done in quantum theory, using the uncertainty principle as an essential element of the argument. Let us consider free elementary particles in $3+1$ dimensions, whose mass are less than $G^{-1} = \kappa$. The motion

of the particles will be linear, at least in some classes of coordinates systems, not accelerating with respect to the natural inertial coordinates at infinity. Let us consider the particle as described by an inertial observer who travels perpendicular to the plane of its motion, which we will call the z direction. From the point of view of that observer, the particles are in an eigenstate of longitudinal momentum, $\hat{P}_z^{\rm tot}$, with some eigenvalue P_z. Since the particles are in an eigenstate of $\hat{P}_z^{\rm tot}$ their wavefunction will be uniform in z, with wavelength L where (note that we assume here that L is so large that we can trust the standard uncertainty relation; besides this uncertainty relation is not being modified in some formulations of DSR)

$$L = \frac{1}{P_z^{\rm tot}} \tag{23}$$

At the same time, we assume that the uncertainties in the transverse positions are bounded a scale r, such that $r \ll 2L$. Then the wavefunction for the the particles has support on a narrow cylinder of radius r which extend uniformly in the z direction. Finally, we assume that the state of the gravitational field is semiclassical, so that to a good approximation, within $\mathcal{C}$ the semiclassical Einstein equations hold.

Since the wavefunction is uniform in z, this implies that the gravitational field seen by our observer will have a spacelike Killing field $k^a = (\partial/\partial z)^a$.

Thus, if there are no forces other than the gravitational field, the semiclassical particles must be described by an equivalent $2+1$ dimensional problem in which the gravitational field is dimensionally reduced along the z direction so that the particles, which are the source of the gravitational field, are replaced by punctures.

The dimensional reduction is governed by a length d, which is the extent in z that the system extends. We cannot take $d < L$ without violating the uncertainty principle. It is then convenient to take $d = L$. Further, since the system consists of the particles, with no intrinsic extent, there is no other scale associated with their extent in the z direction. We can then identify $z = 0$ and $z = L$ to make an equivalent toroidal system, and then dimensionally reduce along z. The relationship between the four dimensional Newton's constant G^4 and the three dimensional Newton's constant $G^3 = G$ is given by

$$G^3 = \frac{G^4}{L} = \frac{G^4 P_z^{\rm tot}}{\hbar} \tag{24}$$

Thus, in the analogous 3 dimensional system, which is equivalent to the original system as seen from the point of view of the boosted observer, the Newton's constant depends on the longitudinal momentum.

Of course, in general there will be an additional scalar field, corresponding to the dynamical degrees of freedom of the gravitational field. However, since we are interested only in the four-dimensional limit, in which local degrees of freedom of the gravitational field are not present, all these fields will vanish this limit.

Now we note that, if there are no other particles or excited degrees of freedom, the energy of the system can to a good approximation be described by

the hamiltonian H of the two dimensional dimensionally reduced system. This is described by a boundary integral, which may be taken over any circle that encloses the particle. But it is well known that in $3d$ gravity H is bounded from above. This may seem strange, but it is easy to see that it has a natural four dimensional interpretation.

The bound is given by

$$M < \frac{1}{4G^3} = \frac{L}{4G^4} \tag{25}$$

where M is the value of the ADM hamiltonian, H. But this just implies that

$$L > 4G^4 M = 2R_{\rm Sch} \tag{26}$$

i.e. this has to be true, otherwise the dynamics of the gravitational field in $3+1$ dimensions would have collapsed the system to a black hole! Thus, we see that the total bound from above of the energy in $2+1$ dimensions is necessary so that one cannot violate the condition in $3+1$ dimensions that a system be larger than its Schwarzschild radius.

Note that we also must have

$$M > P_z^{\rm tot} = \frac{\hbar}{L} \tag{27}$$

Together with (26) this implies $L > l_{\rm Planck}$, which is of course necessary if the semiclassical argument we are giving is to hold.

Now, we have put no restriction on any components of momentum or position in the transverse directions. So the system still has symmetries in the transverse directions. Furthermore, the argument extends to any number of particles, so long as their relative momenta are coplanar. Thus, we learn the following.

Let $\mathcal{H}^{\rm QG}$ be the full Hilbert space of the quantum theory of gravity, coupled to some appropriate matter fields, with $\Lambda = 0$. Let us consider a subspace of states $\mathcal{H}^{\rm weak}$ which are relevant in the low energy limit in which all energies are small in Planck units. We expect that this will have a symmetry algebra which is related to the Poincaré algebra $\mathcal{P}^4$ in 4 dimensions, by some possible small deformations parameterized by G^4 and $\hbar$. Let us call this low energy symmetry group $\mathcal{P}^4_G$.

Let us now consider the subspace of $\mathcal{H}^{\rm weak}$ which is described by the system we have just constructed . It contains the particle, and is an eigenstate of $\hat{P}_z^{\rm tot}$ with large $P_z^{\rm tot}$ and vanishing longitudinal momentum. Let us call this subspace of Hilbert space $\mathcal{H}_{P_z}$.

The conditions that define this subspace break the generators of the (possibly modified) Poincaré algebra that involve the z direction.[7] But they leave unbroken the symmetry in the $2+1$ dimensional transverse space. Thus, a subgroup of $\mathcal{P}^{3+1}_G$ acts on this space, which we will call $\mathcal{P}^{2+1}_G \subset \mathcal{P}^{3+1}_G$.

[7] Notice that if we assume that the four-dimensional rotational symmetry is neither broken, nor deformed, we can recover the whole 4d deformed Poincaré algebra from the 3d one.

We have argued that the physics in $\mathcal{H}_{P_z}$ is to good approximation described by an analogue system of a particle in 2+1 gravity. However, as we will see in the next section the symmetry algebra acting there is not the ordinary 3 dimensional Poincaré algebra, but the κ-Poincaré algebra in 3 dimensions, with

$$\kappa^{-1} = \frac{4G^4 P_z^{\rm tot}}{\hbar} \tag{28}$$

Now we can note the following. Whatever $\mathcal{P}_G^4$ is, it must have the following properties:

- It depends on G^4 and $\hbar$, so that it's action on *each* subspace $\mathcal{H}_{P_z}$, for each choice of P_z, is the κ deformed $3d$ Poincaré algebra, with κ as above.
- It does not satisfy the rule that momenta and energy add, on all states in $\mathcal{H}$, since they are not satisfied in these subspaces.
- Therefore, whatever $\mathcal{P}_G^4$ is, it is not the classical Poincaré group.

Let us therefore turn to gravity coupled with point particle in 2+1 dimension.

4 DSR from 2+1 Dimensional Gravity

Even if not for the argument given in the preceding section, the 2+1 dimensional gravity coupled with point particles would be a perfect test ground for investigating properties of DSR theories. As it is well known this theory is topological, i.e., does not posses any local degrees of freedom, moreover its action

$$S = \int d^3x Tr\left(\mathbf{e} \wedge \mathbf{F}(\boldsymbol{\omega})\right)$$

resembles very closely the four dimensional action of the constrained BF theory in the DSR limit

$$S = \int d^4x Tr\left(\mathbf{B} \wedge \mathbf{F}(\mathbf{A})\right)$$

Investigations in 2+1-dimensional gravity have been pioneered by Staruszkiewicz in 1963 [27], with interest revived by seminal papers by Deser, Jackiw and 't Hooft [28] and Witten [29]. Here we will follow the approach proposed by Matschull and Welling in [30].

The action for (2+1) gravity reads:

$$S = \frac{1}{16\pi G}\int_M d^3x \epsilon^{\mu\nu\rho} Tr(\mathsf{e}_\mu \mathsf{F}_{\nu\rho}) \tag{29}$$

and the basic fields are dreibein e_μ and antisymmetric spin connection ω_μ, whereas G is the gravitational constant, which in (2+1) gravity has a dimension of inverse mass. The Lorentz group in 2+1 dimension $\mathsf{SO}(2,1)$ is isomorphic to $\mathsf{SL}(2,R)$ (which we will in the following denote just $\mathsf{SL}(2)$) and thus the field strength $\mathsf{F}_{\nu\rho}$ defined as:

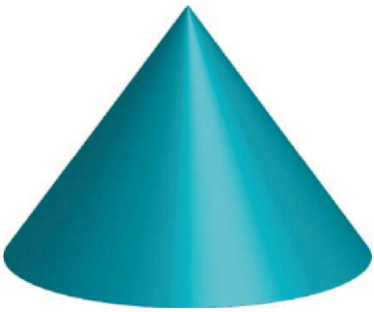


Fig. 1. Space-time of a single particle in 2+1 gravity has conical singularity at the top of the cone ($r = 0$), being the position of the particle. The opening angle of the cone is related to the particle mass and equals $\alpha = 8\pi Gm$. In the case of a spinning particle the space-time geometry is called a spinning cone characterized by deficit angle, related to the mass and time offset related to the spin, see [28]

$$\mathsf{F}_{\mu\nu} = \partial_\mu \omega_\nu - \partial_\nu \omega_\mu + [\omega_\mu, \omega_\nu] \tag{30}$$

is Lie algebra $\mathsf{sl}(2)$-valued. It is convenient to assume that the dreibein e_μ is also $\mathsf{sl}(2)$-valued, where this time the algebra is regarded as a vector space, isomorphic to the three dimensional Minkowski space. As a basis of the $\mathsf{sl}(2)$ algebra we take three dimensional Dirac matrices in real (Majorana) representation and the trace in (29) is just the matrix trace. The field equation following from (29) are

$$\epsilon^{\mu\nu\rho} D_\nu \mathsf{e}_\rho = 0, \quad \epsilon^{\mu\nu\rho} \mathsf{F}_{\nu\rho} = 0 \tag{31}$$

The first equation implies that connection is torsion free, while the second assures the metric is flat. The general solutions of these equations on a simply connected region $U \subset M$ is well known. It consists of the pair of scalar fields (g, f), valued in the Lie group $\mathsf{SL}(2)$ and Lie algebra $\mathsf{sl}(2)$, respectively, such that

$$\omega_\mu = \mathsf{g}^{-1} \partial_\mu \mathsf{g}, \quad \mathsf{e}_\mu = \mathsf{g}^{-1} \partial_\mu \mathsf{f}\, \mathsf{g} \tag{32}$$

where $\mathbf{g} \in SL(2)$ and $\mathbf{f} \in \mathsf{sl}(2)$. Note the similarity between this solution and the solution of the BF theory (10).

Introduction of a particle causes the spacetime to assume the shape of a cone with a particle placed at its top (see Fig. 1). The cone is characterized by the mass-dependent deficit angle α:

$$\alpha = 8\pi Gm \tag{33}$$

where m-particle mass, G-gravitational constant. In what follows we will set $8\pi G = 1$, so that the allowed range of the mass is $m \in [0, \pi)$.

In polar coordinates (t, r, φ), the solution of Einstein equations corresponding to a single spinning particle is of the form

$$\begin{aligned}
\mathbf{e}^0 &= dt + \frac{s}{2\pi}\, d\phi \\
\mathbf{e}^1 &= \left(1 - \frac{m}{2\pi}\right) \cos\phi\, dr - r \sin\phi\, d\phi \\
\mathbf{e}^2 &= \left(1 - \frac{m}{2\pi}\right) \sin\phi\, dr + r \cos\phi\, d\phi \\
\boldsymbol{\omega}^0 &= \frac{m}{2\pi}\, d\phi, \quad \boldsymbol{\omega}^1 = \boldsymbol{\omega}^2 = 0
\end{aligned} \tag{34}$$

This solution corresponds to particle described, similarly to (22) as a delta-like singularity with an appropriate Poincaré charge. There is, however another, more convenient way of treating particles proposed by Matschull and Welling [30], illustrated in Fig. 2. As a result we get singularity free, simply connected spacetime, with boundaries. To make cylindrical boundary look like one dimensional worldline of the particle, we take additional assumption that it circumference vanishes, which can be expressed as requirement that the component $\mathbf{e}_\varphi$ vanishes on the boundary

$$\bar{\mathsf{e}}_\varphi = 0 \text{ at } r = 0\,, \tag{35}$$

where bar marks the value of the field on the boundary.

Since our manifold is simply connected now, a general solutions of Einstein equations in the neighborhood of the boundary is provided by two functions $(\mathsf{f}(r,\varphi,t), \mathsf{g}(r,\varphi,t))$ satisfying (32). On this solution we must impose appropriate boundary conditions, one of which would be (35) at $r = 0$, and another that guarantees continuity of dreibein and connection along the cut $r \geq 0$, $\phi = 0, 2\pi$. For convenience let us denote $\mathsf{f}_\pm(r,t) = \mathsf{f}(r,t;\phi = 0/2\pi)$ and $\mathsf{g}_\pm = \mathsf{g}(r,t;\phi =$

Fig. 2. A different way of description of a point particle has been proposed by Matschull and Welling [30]. Instead of a cone with singularity, they propose to use a manifold with boundaries, constructed as follows. We take the cone in Fig. 1, cut off its tip, and then cut the resulting surface along the line $\phi = 0 = 2\pi$. The resulting, simply connected manifold with boundaries is shown in the figure. On this surface we introduce polar coordinates: $0 \leq r < \infty$, where the line $r = 0$ corresponds to the horizontal boundary, and $0 \leq \phi \leq 2\pi$, where the lines $\phi = 0$ and $\phi = 2\pi$ correspond to the left and right vertical boundaries. To make the surface $r = 0$ looking as a worldline of a particle, we further assume that $e_\phi(r = 0, t) = 0$

$0/2\pi$). Since ω_μ and e_μ are to be continuous at the boundary, $\mathsf{f}_\pm$ and $\mathsf{g}_\pm$ are not independent, and are related by global Poincaré transformation of the form

$$\mathsf{g}_+ = \mathsf{U}^{-1}\mathsf{g}_-, \quad \mathsf{f}_+ = \mathsf{U}^{-1}(\mathsf{f}_- - \mathsf{v})\mathsf{U} \tag{36}$$

where $\mathsf{U} \in \mathsf{SL}(2)$ and $\mathsf{v} \in \mathsf{sl}(2)$ are constant.

Now the condition (35) along with (32)

$$\bar{\mathsf{e}}_\varphi = \bar{\mathsf{g}}^{-1}\partial_\varphi \bar{\mathsf{f}}\bar{\mathsf{g}} = 0 \tag{37}$$

tells that $\bar{\mathsf{f}} = \bar{\mathsf{f}}(t)$. This and the fact that boundary represents worldline the particle makes it possible to identify $\bar{\mathsf{f}}$ with the location of the particle in spacetime

$$\bar{\mathsf{f}}(t, \varphi) = \mathsf{x}(t) \tag{38}$$

Moreover, using the condition $\mathsf{f}_+(t) = \mathbf{f}_-(t)$ and Poincaré transformation, we find that

$$\mathsf{v} = \mathsf{x}(t) - \mathsf{U}\mathsf{x}(t)\mathsf{U}^{-1} \tag{39}$$

Taking time derivative of this equation and making use of the fact that U and v are constants gives

$$0 = \dot{\mathsf{x}}(t) - \mathsf{U}\dot{\mathsf{x}}\mathsf{U}^{-1} \tag{40}$$

This last equation is satisfied if and only if the group element U is of the form

$$\mathsf{U} = u\mathbf{1} + p_a\gamma^a, \quad p_a\gamma^a = \frac{1}{m}\dot{\mathsf{x}} \tag{41}$$

It is natural then to identify p_a with components of momentum of the particle. Note however that this momentum has an unusual property, namely it is, geometrically, a point of the three dimensional anti de Sitter space. Indeed, since $\mathsf{U} \in \mathsf{SL}(2)$, $\det \mathsf{U} = 1$ and thus it follows from (41) that

$$u^2 + p_0^2 - \mathbf{p}^2 = 1 \tag{42}$$

which is just a definition of anti de Sitter space. We see therefore that three dimensional gravity coupled to point particle possesses a fundamental DSR characteristics: the energy-momentum manifold is curved.[8]

It can be shown that instead of the standard dispersion relation the particle on shell satisfies the deformed equation

$$p^a p_a - \sin^2 m = 0 \tag{43}$$

[8] Four dimensional DSR theories with energy-momentum manifolds of the form of anti de Sitter space have not been intensively investigated, though they are known to exist, see [19]. It is not clear if they arise naturally as a limit of 3+1 gravity. It is also not completely clear if de Sitter energy momentum spaces, intensively investigated in the context of DSR in four dimensions, can be obtained in the 2+1 dimensional case.

Of course, in the limit when $m \ll 1$ (remember that the mass scale is set equal 1) we recover the standard dispersion relation.

We know from (32) that the gravitational field in the bulk is pure gauge. It follows that when particles are present the only dynamical degrees of freedom may be associated with boundaries. Therefore, if we start with the action for gravity on the manifold with boundaries, like that in Fig. 2, and then perform symplectic reduction, as the result we find action defined only on the worldline on the particle. As we will see in the moment this action differs from the free particle one: the presence of gravitational field causes deformation of the particle lagrangian, exactly in the DSR spirit.

The procedure described briefly above has been performed by Matschull and Welling [30] and the resulting action reads

$$L = -\frac{1}{2} Tr(\mathsf{U}^{-1}\dot{\mathsf{U}}\,\mathsf{x}) - \varsigma\left(\frac{1}{2}Tr(\mathsf{U}) - \cos m\right) \tag{44}$$

where ς is the Lagrange multiplier enforcing the mass shell constraint (43). Using the expression (41) and the fact that $Tr(\gamma^a\gamma^b) = 2\eta^{ab}$ we can rewrite the Lagrangian in the component form as follows

$$L = -\left(\sqrt{p^2+1}\,\eta^{ab} + \epsilon^{abc}\,p_c - \frac{p^a\,p^b}{\sqrt{p^2+1}}\right) x_b\,\dot{p}_a - \varsigma\left(p^2 - \sin^2 m\right) \tag{45}$$

It can be shown that, in spite of the complex, nonlinear form of the Lagrangian, the resulting equations of motion are just the standard one, to wit

$$\dot{p}_a = 0, \quad \dot{x}^a = \varsigma\, p^a. \tag{46}$$

Let us now turn to discussion of the symmetries of the particle action

$$S = \int d\tau\, L \tag{47}$$

It is clear from the form of the Langrangian (45) that the action is invariant under standard action of Lorentz generators, so that the Lorentz transformations of both position and momentum are the same as in Special Relativity.

To find generators of Lorentz transformations, we should first derive the form of Poisson brackets, resulting from the symplectic potential in (45). These brackets read

$$\{p_a, p_b\} = 0 \tag{48}$$

$$\{p_a, x^b\} = \delta_a{}^b\,\sqrt{p^2+1} + \epsilon_a{}^{bc}\,p_c \tag{49}$$

and

$$\{x^a, x^b\} = 2\epsilon^{ab}{}_c\,x^c \tag{50}$$

Note that this last bracket tells that the positions of the particle do not commute. The exact form of this bracket differs from κ-Minkowski type of noncommutativity (7), but is very closely related to it [26].

Using these bracket it is not difficult to derive the form of Noether charges J_a, generating Lorentz transformations through Poisson bracket. These charges have the form

$$J_a = \sqrt{p^2 + 1}\, \epsilon_{abc}\, x^b\, p^c + 2x_{[a}\, p_{b]}\, p^b \tag{51}$$

and together with conserved momenta they form the standard Poincaré algebra.

Note however that while the action of Lorentz generators J_a on space-time variables x^a is purely classical, the action of translations, generated by momenta p_a is deformed, as a result of the bracket (49). This means that in spite of the fact that the particle lives just in Minkowski space-time the translational invariance is lost (or being deformed). This reminds somehow the model considered by Wess [31], in which also the isometry group is not deformed by itself, but by its action on space-time (or, more generally – phase space) variables.

The form of the particle Lagrangian (44) suggest simple generalization to the case when the energy momentum space is more general than the $\mathsf{SL}(2)$ manifold considered by Matschull and Welling. Consider, for example, the case when this space has the form of de Sitter space. It follows from Iwasawa decomposition of $\mathsf{SO}(d,1)$ group (where d is dimension of space-time and momentum space) that in this case the relevant group element has the form [32], [18]

$$\mathsf{U} = \exp(p_0 t_0) \exp(p_i t_i) \tag{52}$$

where the generators of the "translational" part of Lie algebra $\mathsf{so}(d,1)$ t_0, t_i satisfy the commutational relations reminding the ones of κ-Minkowski space-time

$$[t_0, t_i] = -t_i, \quad [t_i, t_j] = 0 \tag{53}$$

The kinetic term of the Lagrangian reads in this case

$$L_k = -Tr(\mathsf{U}^{-1}\dot{\mathsf{U}}\,\mathsf{x}) = -\left(x^0 - p_i x^i\right)\dot{p}_0 - x^i \dot{p}_i \tag{54}$$

and is invariant under action of the Lorentz group of the form of (3), appended by an appropriate action on position variables (see [17], [10] for details.) In order to get the complete lagrangian, one should add to (54) the term $\varsigma\mathcal{C}$, where $\mathcal{C}$ is the Casimir (3). It can be then checked that κ-Minkowski type of non-commutativity (7) follows from the Lagrangian (54). It is not clear, however, if this Lagrangian can be obtained from gravity directly. Work on this question is in progress.

5 Conclusions

"Seventy-five thousand generations ago, our ancestors set this program in motion," the second man said, "and in all that time we will be the first to hear the computer speak."

"We are the ones who will hear," said Phouchg, "the answer to the great question of Life!.."

"The Universe!.." said Loonquawl.

"And Everything!.."

"Alright," said Deep Thought. "The Answer to the Great Question. . . "

"Yes!.."
"Of Life, the Universe and Everything..." said Deep Thought.
"Yes!.."
"Is..." said Deep Thought, and paused.
"Yes!.."
"Is..."
"Yes!!!?.."
"Forty-two," said Deep Thought, with infinite majesty and calm.[9]

The current status of DSR reminds somehow the Adams' "forty-two", the answer to the question, which we do not really know. To be honest, we do not have any proof yet that this answer is correct, though we hope that Pierre Auger Observatory and GLAST satellite will provide us with such a proof. However, as we tried to argue above, there are more and more indications that the right question is "What is the semiclassical, flat space limit of quantum gravity?" It is our hope that it would not require seventy-five thousand generations to convince ourself that this hypothesis is correct.

Acknowledgements

For JKG this work is partially supported by the KBN grant 1 P03B 01828.

References

1. D. Colladay and V.A. Kostelecky, "Lorentz-Violating Extension of the Standard Model," Phys. Rev. **D 58**, 116002 (1998) [arXiv:hep-ph/9809521]; R. Bluhm, V.A. Kostelecky, and N. Russell, "CPT and Lorentz Tests in Hydrogen and Antihydrogen," Phys. Rev. Lett. **82**, 2254 (1999) [arXiv:hep-ph/9810269]; R. Bluhm and V.A. Kostelecky, "Lorentz and CPT Tests in Spin-Polarized Solids," Phys. Rev. Lett. **84**, 1381 (2000) [arXiv:hep-ph/9912542]; V.A. Kostelecky and M. Mewes; "Signals for Lorentz Violation in Electrodynamics," Phys. Rev. **D 66**, 056005 (2002) [arXiv:hep-ph/0205211]; R. Bluhm, V.A. Kostelecky, C.D. Lane, and N. Russell, "Probing Lorentz and CPT Violation with Space-Based Experiments," Phys. Rev. **D 68** (2003) 125008 [arXiv:hep-ph/0306190]; R. C. Myers and M. Pospelov, "Ultraviolet modifications of dispersion relations in effective field theory," Phys. Rev. Lett. **90** (2003) 211601 [arXiv:hep-ph/0301124]; see also Robert Bluhm contribution to this volume.
2. J. Alfaro, H. A. Morales-Tecotl and L. F. Urrutia, "Quantum gravity corrections to neutrino propagation," Phys. Rev. Lett. **84** (2000) 2318 [arXiv:gr-qc/9909079]; J. Alfaro, H. A. Morales-Tecotl and L. F. Urrutia, "Loop quantum gravity and light propagation," Phys. Rev. D **65** (2002) 103509 [arXiv:hep-th/0108061]; J. Alfaro, M. Reyes, H. A. Morales-Tecotl and L. F. Urrutia, "On alternative approaches to Lorentz violation in loop quantum gravity inspired models," Phys. Rev. D **70** (2004) 084002 [arXiv:gr-qc/0404113]; see also Luis Urrutia contribution to this volume.

[9] Douglas Adams. The Hitch Hikers Guide to Galaxy Fantazy. 1990.

3. T. Jacobson, S. Liberati and D. Mattingly, "Astrophysical bounds on Planck suppressed Lorentz violation," arXiv:hep-ph/0407370.
4. T. Jacobson, S. Liberati and D. Mattingly, "Lorentz violation at high energy: concepts, phenomena and astrophysical constraints," arXiv:astro-ph/0505267.
5. L. Smolin, "Falsifiable predictions from semiclassical quantum gravity," arXiv:hep-th/0501091.
6. G. Amelino-Camelia, "Testable scenario for relativity with minimum-length," Phys. Lett. B **510**, 255 (2001) [arXiv:hep-th/0012238].
7. G. Amelino-Camelia, "Relativity in space-times with short-distance structure governed by an observer-independent (Planckian) length scale," Int. J. Mod. Phys. D **11**, 35 (2002) [arXiv:gr-qc/0012051].
8. J. Kowalski-Glikman, "Observer independent quantum of mass," Phys. Lett. A **286** (2001) 391 [arXiv:hep-th/0102098].
9. N. R. Bruno, G. Amelino-Camelia and J. Kowalski-Glikman, "Deformed boost transformations that saturate at the Planck scale," Phys. Lett. B **522** (2001) 133 [arXiv:hep-th/0107039].
10. J. Kowalski-Glikman, "Introduction to doubly special relativity," in Giovanni Amelino-Camelia and Jerzy Kowalski-Glikman, *Planck Scale Effects in Astrophysics and Cosmology*, Lecture Notes in Physics **669**, Springer 2005 [arXiv:hep-th/0405273].
11. J. Lukierski, H. Ruegg, A. Nowicki and V. N. Tolstoi, "Q deformation of Poincare algebra," Phys. Lett. B **264** (1991) 331.
12. J. Lukierski, A. Nowicki and H. Ruegg, "Real forms of complex quantum anti-De Sitter algebra U_q(Sp(4:C)) and their contraction schemes," Phys. Lett. B **271** (1991) 321 [arXiv:hep-th/9108018].
13. S. Majid and H. Ruegg, "Bicrossproduct structure of kappa Poincare group and noncommutative geometry," Phys. Lett. B **334** (1994) 348 [arXiv:hep-th/9405107].
14. J. Lukierski, H. Ruegg and W. J. Zakrzewski, "Classical quantum mechanics of free kappa relativistic systems," Annals Phys. **243** (1995) 90 [arXiv:hep-th/9312153].
15. J. Magueijo and L. Smolin, "Lorentz invariance with an invariant energy scale," Phys. Rev. Lett. **88** (2002) 190403 [arXiv:hep-th/0112090].
16. J. Magueijo and L. Smolin, "Generalized Lorentz invariance with an invariant energy scale," Phys. Rev. D **67** (2003) 044017 [arXiv:gr-qc/0207085].
17. J. Kowalski-Glikman and S. Nowak, "Doubly special relativity and de Sitter space," Class. Quant. Grav. **20** (2003) 4799 [arXiv:hep-th/0304101].
18. J. Kowalski-Glikman and S. Nowak, "Quantum kappa-Poincare algebra from de Sitter space of momenta," arXiv:hep-th/0411154.
19. A. Blaut, M. Daszkiewicz, J. Kowalski-Glikman and S. Nowak, "Phase spaces of doubly special relativity," Phys. Lett. B **582** (2004) 82 [arXiv:hep-th/0312045].
20. L. Freidel and A. Starodubtsev, "Quantum gravity in terms of topological observables," arXiv:hep-th/0501191.
21. S. W. MacDowell and F. Mansouri, "Unified Geometric Theory Of Gravity And Supergravity," Phys. Rev. Lett. **38**, 739 (1977) [Erratum-ibid. **38**, 1376 (1977)].
22. L. Freidel, K. Krasnov and R. Puzio, "BF description of higher-dimensional gravity theories," Adv. Theor. Math. Phys. **3**, 1289 (1999) [arXiv:hep-th/9901069].
23. L. Smolin and A. Starodubtsev, "General relativity with a topological phase: An action principle," arXiv:hep-th/0311163.
24. S. Alexander, "A quantum gravitational relaxation of the cosmological constant," arXiv:hep-th/0503146.

25. A. P. Balachandran, G. Marmo, B. S. Skagerstam and A. Stern, "Gauge Symmetries And Fiber Bundles: Applications To Particle Dynamics," Lect. Notes Phys. **188** (1983) 1.
26. L. Freidel, J. Kowalski-Glikman and L. Smolin, "2+1 gravity and doubly special relativity," Phys. Rev. D **69** (2004) 044001 [arXiv:hep-th/0307085].
27. A. Staruszkiewicz, Acta Phys. Polon. **24**, 734 (1963).
28. S. Deser, R. Jackiw and G. 't Hooft, "Three-Dimensional Einstein Gravity: Dynamics Of Flat Space," Annals Phys. **152**, 220 (1984).
29. E. Witten, "(2+1)-Dimensional Gravity As An Exactly Soluble System," Nucl. Phys. B **311**, 46 (1988).
30. H. J. Matschull and M. Welling, "Quantum mechanics of a point particle in 2+1 dimensional gravity," Class. Quant. Grav. **15** (1998) 2981 [arXiv:gr-qc/9708054].
31. J. Wess, "Deformed coordinate spaces: Derivatives," arXiv:hep-th/0408080.
32. J. Kowalski-Glikman, "De Sitter space as an arena for doubly special relativity," Phys. Lett. B **547** (2002) 291 [arXiv:hep-th/0207279].

Corrections to Flat-Space Particle Dynamics Arising from Space Granularity

L.F. Urrutia

Instituto de Ciencias Nucleares, Universidad Nacional Autónoma de México, Circuito Exterior, C.U., 04510 México, D.F.
urrutia@nucleares.unam.mx

Abstract. The construction of effective Hamiltonians describing corrections to flat space particle dynamics arising from the granularity of space at very short distances is discussed in the framework of an heuristic approach to the semiclassical limit of loop quantum gravity. After some general motivation of the subject, a brief non-specialist introduction to the basic tools employed in the loop approach is presented. The heuristical semiclassical limit is subsequently defined and the application to the case of photons and spin 1/2 fermions is described. The resulting modified Maxwell and Dirac Hamiltonians, leading in particular to Planck scale corrections in the energy-momentum relations, are presented. Alternative interpretations of the results and their limitations, together with other approaches are briefly discussed along the text. Three topics related to the above methods are reviewed: (1) The determination of bounds to the Lorentz violating parameters in the fermionic sector, obtained from clock comparison experiments. (2) The calculation of radiative corrections in preferred frames associated to space granularity in the framework of a Yukawa model for the interactions and (3) The calculation of synchrotron radiation in the framework of the Myers–Pospelov effective theories describing Lorentz invariance violations, as well as a generalized approach to radiation in Planck scale modified electrodynamics. The above exploratory results show that quantum gravity phenomenology provides observational guidance in the construction of quantum gravity theories and opens up the possibility of probing Planck scale physics.

1 Introduction

Most theories of gravity suggest that our notion of space-time as a continuum needs to be revised at short distances (high energies) of the order of the Planck length $\ell_P \approx 10^{-33}$ cm (Planck mass, $M_P \approx 10^{19}$ GeV). At these scales quantum effects should be important and since space-time is to be considered as a set of dynamical interacting variables, instead of a mere background where physics occurs, the quantum nature of them together with their corresponding fluctuations could induce modifications to our standard notion of space-time. The consequences of space-time been considered as a continuum have been already successfully probed up to the much lower energies of $\approx 10^3$ GeV, corresponding

L.F. Urrutia: *Corrections to Flat-Space Particle Dynamics Arising from Space Granularity*, Lect. Notes Phys. **702**, 299–345 (2006)
DOI 10.1007/3-540-34523-X_11
© Springer-Verlag Berlin Heidelberg 2006

to the standard model of particle physics. The sixteen orders of magnitude between our current experimental arena and the region where such new effects would become relevant provides in fact plenty of room to look for modifications of our generally accepted ideas of space-time. We will refer to this new possibility as space-time presenting a granular or foamy structure in the following. Central to this question is the long time honored problem of finding a consistent unification of gravity and quantum mechanics, which nowadays is been actively pursued by loop quantum gravity [1] and string theory [2], among others lines of research. At any rate, a complete formulation of quantum gravity must explain how the standard notion of space-time at macroscopical length scales is recovered, thus validating the many successful test of classical Einstein gravity that have been performed and which have served to make sure that the correct starting point to construct the full theory has been taken.

Then a reasonable question to ask is whether or not such short length (high energy) effects leave any imprint in the dynamics of particles at standard model energies, which we could be able to detect with present day technology and observational sensibilities.

On one hand, from a purely phenomenological point of view one would expect some modifications to arise in the same way as particle propagation properties change when they move in a medium, with respect to those in the vacuum. Of course the analogy of a modified structure of space-time with a propagating medium is at most very tentative because the idea of a medium presupposes the existence of something external in which it is embedded. On the contrary, space-time is the arena where phenomena occur and there is nothing external to it. Nevertheless it is very plausible that a drastic change in our description of space-time would induce modifications in the way we deal with physics at such very short scales. Here we take the point of view that there is a remnant of such modifications at standard model energies, which are described by an effective field theory valid up to scales much lower than the Planck mass. These corrections manifest themselves as additional terms contributing to the propagation and interactions of known particles.

On the other hand, for a long time it has been taken for granted that Planck scale phenomena is completely out of reach from present experiments and/or observations (astrophysical, for example). Recent investigations show that this is definitely not the case, thus opening the door to a new area of research called Quantum Gravity Phenomenology (QGP) that is designed to use existing and forthcoming experiments and/or observations to restrict or constraint competing theories of quantum gravity based on their predicted imprints at these lower energies [3]. A partial list of references describing such efforts is given in [4, 5]. This is definitely a great advance over the purely aesthetically criteria that prevailed before. Even though QGP has made independent advances in restricting the different parameters encoding some of the proposed modifications induced at standard model energies, there still remain the open problem of calculating them as a rigorous semiclassical approximation of a fundamental theory. Only after this gap is filled one could really be in position to use the observations

to restrict the latter. The work presented here summarizes an heuristical step towards the estimation of the flat space semiclassical limit in the realm of LQG, together with further elaborations upon the results obtained in this way. Also some closely related topics originating from different approaches are discussed.

Before going to the details, let us briefly describe some of the alternative points of view to be found in the literature regarding the question of whether or not a granular or foamy structure of space would induce modifications to particle dynamics at standard model energies. Such modifications have been mainly understood in terms of the violation of standard Lorentz covariance, either through the introduction of preferred reference frames or by means of the inclusion of an extended or deformed Lorentz relativity. In the following we will generically refer to theses possibility as Lorentz invariance violation (LIV).

A first viewpoint is that modifications that manifest themselves in the form of LIV, do not arise. In this way covariance under the standard Lorentz transformations would be perfectly compatible with a discrete nature of space at Planck scale, which is encoded in the discrete spectrum of area and volume operators in LGQ, for example. This would be analogous to the well known property that the discrete spectrum of the angular momentum operator does not preclude the invariance of a system under the continuous rotation group [6, 7].

A second possibility is that corrections in the form of LIV do in fact arise. Within this point of view different alternatives have been also considered, which can be separated in two further categories:

(i) The first one can be characterized by a phenomenological parameterization of all possible corrections terms, according to the dimensionality of the corresponding LIV operators, which are assumed to arise via a spontaneous Lorentz symmetry breaking of a more fundamental model like string theory, for example [8]. These vacuum expectation values define a set of preferred frames, called concordant frames [9], in which the LIV terms can be maintained appropriately small when going from one frame to another via a passive (observer) Lorentz transformation. This is analogous to the description of atomic phenomena in the presence of an external magnetic field, where rotational invariance is broken by active (particle) transformation. Nevertheless, one can rotate the apparatus and perform the experiment in the presence of the rotated external field. The Standard Model Extension (SME) [11] belongs to this class and it has been highly successful in allowing a unified description of the great amount of experimental data that has been gathered since 1960 [10] in relation to the experimental verification of the isotropy of apace, the transformation properties among inertial frames and the validity of the discrete transformations C, P and T as well as its possible combinations [12]. The SME has been recently generalized to incorporate gravity [13]. More recently LIV models based upon dimension five operators have been constructed [14] and thoroughly analyzed in [15, 16]. Within the approach (i) we find also direct extensions of Dirac and Maxwell equations incorporating modifications which go beyond effective field theories [17]. An

alternative approach to LIV via spontaneous symmetry breaking can be found in [18].

(ii) A second broad category of researches which leads to dynamical corrections can be distinguished by an effort to obtain such modifications either from an extension or deformation of the standard Lorentz relativity principle [19–24], or from a fundamental description of quantum gravity like, for example: an effective field theory description of quantum general relativity [25,26], a version of non-critical string theory [27–29] or the approach of loop quantum gravity [30–35]. Alternative ways of incorporating such dynamical modifications can be found in [36,37].

The most immediate way in which the corrections to the dynamics show up is through modifications to the energy-momentum relations for the particles [38,39]

$$\omega_{\pm}^2 \quad = |\mathbf{k}|^2 \; \pm \xi \frac{|\mathbf{k}|^3}{E_{QG}} \; , \tag{1}$$

$$E_{R,L}^2 = |\mathbf{p}|^2 + m^2 + \eta_{R,L} \frac{|\mathbf{p}|^3}{E_{QG}} \; , \tag{2}$$

as shown above for photons and fermions, respectively. The connection of such modifications with gravity, together with the possible astrophysical observation of the energy-dependent velocity in high-energy photons arriving from cosmological sources was suggested in [40]. In fact, the frequency dependent photon velocity

$$|\mathbf{v}_{\pm}| = \frac{\partial \omega}{\partial |\mathbf{k}|} = \left(1 \pm \xi \, \frac{3}{2} \frac{|\mathbf{k}|}{E_{QG}} + \ldots \right) \tag{3}$$

predicts a time delay Δt in the arrival of two photons having an energy difference ΔE and travelling a distance L, given by

$$\Delta t \approx \xi \frac{\Delta E}{E_{QG}} \frac{L}{c} \; . \tag{4}$$

Here E_{QG} denotes the scale where quantum gravity effects become relevant, which is usually taken as the Planck energy. As emphasized in [40], the tiny contribution of ξ / E_{QG} can be amplified by selecting large energy differences or, more effectively, distances L of cosmological magnitude. It is important to emphasize that this time delay is a purely kinematical consequence of the modified dispersion relations. The orders of magnitude $L \approx 10^{10}$ l.y., $\Delta E \approx 20$ MeV, $E_{QG} = 10^{19}$ GeV, $\xi \approx 1$, lead to $\Delta t \approx 10^{-3}$ s, which is within the range of sensitivities δt in actual and forthcoming gamma ray bursts (GRB) observations. In order to measure such effect it is necessary that $\delta t < \Delta t$. Present sensitivities allow for estimates of the lower bounds for E_{QG}/ξ presented in Table 1. Future planned observations of GRB at cosmological distances having well determined red shifts z, together with greater sensitivities ranging form 10^{-6} s (RHESSI: Reuben Ramaty High Energy Solar Spectrometer) to 10^{-7} s (GLAST: Gamma Ray Large Area Telescope) will allow a substantial increase of such bounds. The fireball

Table 1. Bounds upon the quantum gravity scale E_{QG}/ξ

z	$\delta t(s)$	Objects	$E_{QG}/\xi(GeV)$	Source
0.031	280	Markarian421	$> 4.0 \times 10^{16}$	[4]
0.0085-3.42	0.064	GRB′s (BATSE)	$> 6.9 \times 10^{15}$	[41]
0.3	0.031	GRB021206 (RHESSI)	$> 1.3 \times 10^{17}$	[42]

model of GRB emission [43, 44] predicts also the generation of $10^5 - 10^{10}$ GeV neutrino bursts which will be detected by observatories like NUBE (Neutrino Burst Experiment), OWL (Orbiting Wide-angle Light collector experiment) and EUSO (Extreme Universe Space Observatory), for example. This will open up the possibility of using such particles, some times detected in coincidence with the respective photons, to set further bounds upon the quantum gravity scale E_{QG}.

Very soon after the proposal of [40], the first derivation of a consistent electrodynamics leading to the dispersion relations (1) was obtained in the framework of a loop quantum gravity inspired model [30]. The resulting modified Maxwell equations are

$$\nabla \cdot \mathbf{E} = 0, \quad \partial_t \mathbf{E} = -\nabla \times \mathbf{B} + 2\,\xi\,\ell_P\,\nabla^2\,\mathbf{B}\,, \tag{5}$$

$$\nabla \cdot \mathbf{B} = 0, \quad \partial_t \mathbf{B} = +\nabla \times \mathbf{E} - 2\,\xi\,\ell_P\,\nabla^2\,\mathbf{E}\,, \tag{6}$$

leading to the energy-momentum relation

$$\omega_\pm = |\mathbf{k}|\,(1 \mp 2\,\xi\,\ell_P\,|\mathbf{k}|)\ . \tag{7}$$

Additional bounds upon the parameters describing the quantum gravity induced modifications have been obtained by incorporating the dynamics through the SME or similar constructions. In particular, the topics of Lorentz and CPT violations have been thoroughly studied in low energy physics via theory and experiments related to: Penning traps, clock comparison measurements, hydrogen-antihydrogen studies, spin polarized dispersion and muon experiments, among others subjects [45].

Finally we mention the use of astrophysical phenomena to discuss such modified theories. Distinguished examples are the bounds imposed by polarization measurements from astrophysical sources [46], the study of ultra high energy physics processes, among them cosmic rays, [47–51] and the consequences of the detected synchrotron radiation from the Crab nebulae, as well as that from other objects [15, 16, 52, 53]. For recent reviews about such topics see [54].

This contribution summarizes the work carried over in collaborations with different colleagues and it is organized as follows. Section 2 contains a non-specialist review of some basic elements of LQG to be subsequently employed in the estimation of the induced dynamical modifications. In Sect. 3 we describe the general features involved in the heuristical calculation of the effective photon and fermion Hamiltonians incorporating Planck scale corrections induced by LQG.

These Hamiltonians are presented in Subsect. 3.3 together with some related comments. The remaining Sect. 4 is devoted to selected phenomenological applications of models presenting Planck scale dynamical corrections. Using existing data from clock-comparison experiments, in Subsect. 4.1 we obtain stringent bounds upon combinations of parameters appearing in the effective fermionic Hamiltonian previously derived. The incorporation of radiative corrections to the description of LIV effects in preferred frames associated to space granularity is discussed in Subsect. 4.2, leading to severe fine-tuning and naturalness problems. Subsection 4.3 contains the discussion of synchrotron radiation in LIV electrodynamics, emphasizing the model of [14], which is phenomenologically formulated as a theory parameterizing LIV with dimension five operators. Finally, a unified description of radiation in Planck scale modified electrodynamics including different models is presented in Subsect. 4.4.

2 Basic Elements from Loop Quantum Gravity (LQG)

Detailed reviews of LQG can be found in [1, 55]. Here we just indicate some of the basic features that will be relevant in our heuristic calculation and present an intuitive, non-specialist introduction to this approach. LQG turns out to be the formulation of Einstein gravity as a Hamiltonian gauge theory with additional constraints, written in terms of non-local gauge covariant, diffeomorphisms invariant quantities. It is formulated in four dimensions and the matter couplings are obtained by rewriting the standard ones in terms of the new variables. This background independent, non-perturbative theory has allowed substantial progress in the old problem of producing a consistent quantum description of Einstein gravity. It has successfully dealt with traditional problems like a macroscopic account of the black hole entropy [56] and the construction of non-singular cosmological models [57], among other topics. One of its most notable predictions is the property that area and volume operators are quantized in terms of the corresponding powers of ℓ_P, thus signaling a granular structure of space at short distances [58]. It is precisely the possible consequences that such granularity may induce at energy scales of the standard model what is to be explored in this work. This question is intimately related to the still open problem of the semiclassical limit in LQG: how does one recover the continuous metric description of space-time starting from the quantum version of it?.

2.1 The Passage to the New Variables

LQG is formulated in terms of the Ashtekar-Barbero variables that arise as a canonical transformation from the usual ADM variables. Here we summarize the main steps leading to this choice and closely follow [55].

Let us start from the standard Einstein-Hilbert action for gravity in the signature $(-,+,+,+)$

$$S[g_{\mu\nu}] = \frac{1}{2\kappa}\int \left(d^4x\right)\ \sqrt{-\det(g_{\mu\nu})}\ {}^{(4)}R, \quad \kappa = \frac{8\pi G}{c^3}\ , \tag{8}$$

where ${}^{(4)}R$ is the four dimensional Ricci scalar. The ADM variables result from a foliation of spacetime in 3D surfaces Σ of constant parameter t, described by coordinates $x^a, a = 1, 2, 3$, together with the parameterization of the ten variables $g_{\mu\nu}$ in terms of the six components on the induced three-metric q_{ab} on Σ, plus the three components of the shift vector N^a and the lapse function N. In terms of these variables the four dimensional metric is

$$g_{tt} = q_{cd}N^c N^d - N^2, \quad g_{ta} = g_{at} = q_{ac}N^c, \quad g_{ab} = q_{ab} \tag{9}$$

and the action (8) reads

$$S[q_{ab},\ N^a,\ N] = \frac{1}{2\kappa}\int dt \int \left(d^3x\right)\ \sqrt{\det(q_{ab})}\ N\ \left({}^{(3)}R + K_{ab}K^{ab} - (K^a_a)^2\right)\ , \tag{10}$$

where all the indices are now lowered or raised by q_{ab} and its inverse q^{cd}. Here ${}^{(3)}R$ is the Ricci scalar of the manifold Σ. The information about the velocities $\dot{q}_{ab}$ is contained in the extrinsic curvature

$$K_{ab} = \frac{1}{2N}\left(\dot{q}_{ab} - \mathcal{L}_{\mathbf{N}}\ q_{ab}\right) = K_{ba}\ , \tag{11}$$

where $\mathcal{L}_{\mathbf{N}}$ denotes de Lie derivative along the vector N^a. Introducing the canonical momenta

$$\frac{1}{\kappa}\Pi^{ab} = \frac{\delta S}{\delta \dot{q}_{ab}} = \frac{1}{\kappa}\left(K^{ab} - q^{ab}K^c_c\right)\sqrt{\det(q_{ab})}\ , \tag{12}$$

$$\Pi = \frac{\delta S}{\delta \dot{N}} = 0, \qquad \Pi_a = \frac{\delta S}{\delta \dot{N}^a} = 0 \tag{13}$$

and making a Legendre transformation one finds

$$S[q_{ab},\ \Pi^{ab},\ \ N^a,\ N] = \frac{1}{2\kappa}\int dt \int \left(d^3x\right)\left[\Pi^{ab}\,\dot{q}_{ab} - N^a H_a - N\ H\right]\ . \tag{14}$$

The above action shows that gravity is a fully constrained theory with zero canonical Hamiltonian. The invariance under local diffeomorphisms in Σ is generated by the constraints H_a, while the dynamics is generated by the Hamiltonian constraint H. These are first class constraints which can be written as explicit functions of the canonical variables. The shift and the lapse functions turn out to be Lagrange multipliers.

The symplectic structure corresponding to (14) is given by the Poisson brackets

$$\left\{\Pi^{ab}(\mathbf{x}),\ q_{cd}(\mathbf{y})\right\} = 2\kappa\delta^a_{(c}\delta^b_{d)}\delta^{(3)}(\mathbf{x},\mathbf{y})\ , \tag{15}$$

with the remaining ones being zero.

The second step towards the construction of the new variables is to introduce a non-abelian $SO(3) \approx SU(2)$ formulation of the action (14). To this end one rewrites the three-metric in terms of a triad e^i_a such that

$$q_{ab} = e^i_a e^j_b \delta_{ij} \ , \tag{16}$$

where the indices $i, j = 1, 2, 3$, transform under local rotations.

The canonical variables arising in this step are the densitisized (weight +1) triad

$$E^a_i = \frac{1}{2} \epsilon^{abc} \epsilon_{ijk} e^j_b e^k_c \ , \tag{17}$$

together with the projected extrinsic curvature

$$K^i_a = K_{ab} E^b_j \delta^{ij} \ . \tag{18}$$

In terms of them we can rewrite

$$\det(q_{ab}) = \det(E^a_i), \qquad \dot{q}_{ab} \Pi^{ab} = 2 \dot{K}^i_a \, E^a_i \ , \tag{19}$$

and the action is

$$S\left[E^a_i, \ K^j_b, N^a, N, N^i\right] = \frac{1}{\kappa} \int dt \int \left(d^3x\right) \Big[E^a_i \, \dot{K}^i_a - N^a H_a(E,K) \\ - N \, H(E,K) - N^i G_i(E,K)\Big] \ , \tag{20}$$

The introduction of an extra $SO(3)$ gauge freedom in (16) requires the additional constraints

$$G_i(E,K) = \epsilon_{ijk} E^{aj} K^k_a \ , \tag{21}$$

which turn out to be just the non-abelian Gauss law.

The final step to the Ashtekar-Barbero variables [59] starts from the recognition that there is a natural $SO(3) \approx SU(2)$ connection Γ^i_a compatible with the triad, such that

$$\partial_{[a} e^i_{b]} + \epsilon^i{}_{jk} \Gamma^j_{[a} e^k_{b]} = 0 \ . \tag{22}$$

Moreover, the triad introduced in (16) can be expressed in terms of the densitized triad as

$$e^i_a = \frac{1}{2\sqrt{|\det(E)|}} \, \epsilon_{abc} \epsilon^{ijk} E^b_j E^c_k, \qquad e^a_i = \frac{sgn(\det(E))}{\sqrt{|\det(E)|}} \, E^a_i \ . \tag{23}$$

The final Ashtekar-Barbero connection, which allows to write the constraints in a simpler form, is

$$A^i_a = \Gamma^i_a + \gamma K^i_a \ . \tag{24}$$

Here γ is the Immirzi parameter [60] and the corresponding field strength two-form is

$$F^i = dA^i + \epsilon^i{}_{jk} A^j \wedge A^k, \qquad F^{ij}_{ab} = \epsilon^{ij}{}_k \, F^k_{ab} \ . \tag{25}$$

The final action results

$$S\left[E^a_i, \ A^j_b, N^a, N, N^i\right] = \frac{1}{\kappa} \int dt \int \left(d^3x\right) \Big[E^a_i \, \dot{A}^i_a - N^a H_a(E,A) \\ - N \, H(E,A) - N^i G_i(E,A)\Big] \ , \tag{26}$$

with the canonical variables satisfying the non-zero Poisson brackets

$$\{E^b_j(\mathbf{y}), A^i_a(\mathbf{x})\} = \kappa\,\gamma\,\delta^b_a\,\delta^i_j\,\delta^{(3)}(\mathbf{x},\mathbf{y})\ . \tag{27}$$

The explicit form of the constraints are

$$\begin{aligned} G_i &= D_a\ E^a_i, \qquad H_b = E^a_i\,F^i_{ab} - (1+\gamma^2)K^i_b\,G_i, \\ H &= \frac{1}{\sqrt{|\det(E)|}}\left(F^{ij}_{ab} - 2(1+\gamma^2)K^i_{[a}\,K^j_{b]}\right)\,E^a_i\,E^b_j\ . \end{aligned} \tag{28}$$

The choice $\gamma = i$ corresponds to the original Ashtekar variables, which define a complexified version of Einstein gravity together with a very simple structure for the Hamiltonian constraint. The price to be paid is the need to incorporate some reality conditions in order to recover the corresponding real formulation.

Following the standard Dirac quantization procedure arising from the action (26), together with the Poisson brackets (27) in the coordinate space defined by the connection A^i_a, we promote the canonical variables to operators in such a way that

$$\hat{A}^i_a\Psi[A] = \ A^i_a\Psi[A], \qquad \hat{E}^a_i\Psi[A] = -i\hbar\,\kappa\,\gamma\frac{\delta}{\delta A^i_a}\Psi[A] \tag{29}$$

and impose the operator version of the constraints as null conditions upon the wave function.

2.2 Holonomies and Fluxes

Next we describe the fundamental operator variables employed to formulate LQG and which supersede the quantum version of the previous canonical variables (17) and (24). The basic support for the operators in LQG are open or closed curves in Σ, as depicted in Fig. 1, respectively, together with two-dimensional surfaces. These allow to define appropriately smeared versions of the operators (17) and (24). The $SU(2)$ connection operators $A^i_a(\mathbf{x})$ are replaced by holonomies (or parallel transport matrices) U along open or closed curves in the thee-manifold Σ, defined as

$$U(\gamma, A) = \mathcal{P}\exp i\int_{0,\,\gamma}^{s} A^i{}_a(\gamma(s'))\,\sigma_i\,\frac{dx^a}{ds'}\,ds' \quad \in SU(2)\ . \tag{30}$$

The notation is the following: the Pauli matrices σ^i are the generators of the $SU(2)$ Lie algebra. The open curve γ is parameterized by s' with $0 < s' < s$ and

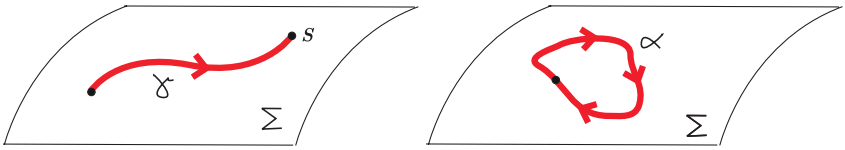


Fig. 1. Open and closed curves in Σ

$\mathcal{P}$ denotes the standard path-ordered product. The holonomy, which is a multiplicative operator, transforms covariantly under $SU(2)$ gauge transformations at the end points, but still it is not invariant under diffeomorphisms.

The canonically conjugated momenta operators $\hat{E}_i^a$ are replaced by their corresponding fluxes over surfaces S in Σ

$$\hat{E}_i(S) = -i\hbar \int_S d\sigma^1\, d\sigma^2\, n_a(\boldsymbol{\sigma})\, \frac{\delta}{\delta\ A_a^i(x(\boldsymbol{\sigma}))}, \quad n_a = \epsilon_{abc} \frac{\partial x^b}{\partial \sigma^1} \frac{\partial x^c}{\partial \sigma^2}\ . \tag{31}$$

Here $\boldsymbol{\sigma} = (\sigma^1, \sigma^2)$ are the intrinsic coordinates of the surface and n_a is its normal vector. The above definitions (30) and (31) allow to calculate the corresponding commutator, which turns out to be a non-canonical one.

Quantum states $\Psi_{\Gamma,f}(A)$ are represented by functionals of generalized connections ($SU(2)$ group elements) defined over graphs in Σ, which are called cylindrical functions and define the kinematical space of the problem. As represented in Fig. 2, a graph $\Gamma = \{\mathbf{V}_1, \ldots, \mathbf{V}_n; \gamma_1, \ldots, \gamma_m\}$ is a set of points $\{\mathbf{V}_1, \ldots, \mathbf{V}_n\} \in \Sigma$, called the vertices, joined by curves $\gamma_1, \ldots, \gamma_m$, called the edges of the graph. The number of edges attached to a vertex is called the valence of the vertex. To each edge of the graph we associate a group element $U(\gamma, A)$ labelled by an irreducible representation of $SU(2)$ and consider a function f which is a map from the direct product of group representations to the complex numbers, so that

$$\Psi_{\Gamma,f}(A) = f(U(\gamma_1, A), \ldots, U(\gamma_m, A)) \ \in \ \text{Kinematical Space}\ . \tag{32}$$

For a given graph Γ we define the scalar product of two associated functions f and g as

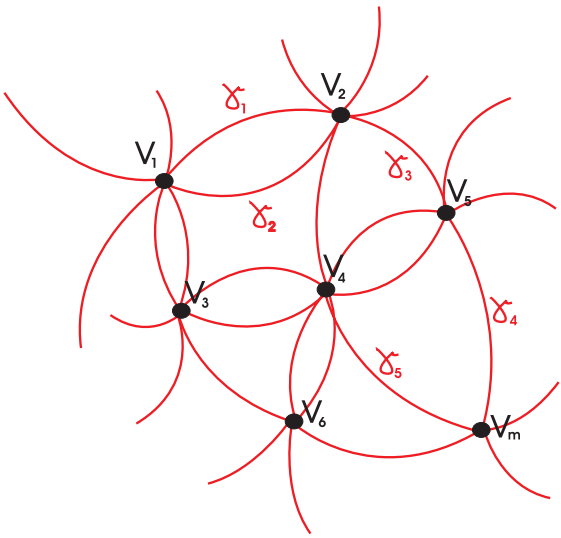


Fig. 2. Piece of a graph $\Gamma = \{\mathbf{V}_1, \ldots, \mathbf{V}_n; \gamma_1, \ldots, \gamma_m\}$

$$\langle \Psi_{\Gamma,f} | \Psi_{\Gamma,g} \rangle = \int \left(\prod_{\gamma_i} dU_i \right) f^*(U_1 \dots U_m) g(U_1 \dots U_m \quad , \tag{33}$$

where we have denoted $U_i := U(\gamma_i, A)$. Here dU_i is the corresponding Haar measure for the group $SU(2)$.

There is still a long way to construct the physical space. To this end one has to subsequently impose $SU(2)$ gauge invariance, three dimensional diffeomorphism invariance and finally the Hamiltonian constraint. The first two conditions can be explicitly accomplished. The first step is implemented by an adequate choice of the functions f in terms of the so-called intertwiners at each vertex. These are invariant tensors in $SU(2)$ which map the product of representations at each vertex into gauge invariant expressions. In these way the so called spin network states are defined and it is possible to prove that they form an orthonormal basis for the $SU(2)$ gauge invariant kinematical space. The second step is performed by going from loops to knots, thus defining equivalent classes of loops under diffeomorphisms.

The action of the operators (30) and (31) upon quantum states (33) is defined by

$$\begin{aligned} (U_B^A(A,\gamma)\Psi)[A] &= U_B^A(A,\gamma) \times \Psi[A] \ , \\ \hat{E}_i(S)U(A,\gamma) \quad &= \sum_k \pm i\hbar\, U(A,\gamma_1^{P_k})\, \sigma_i\, U(A,\gamma_2^{P_k}) \ . \end{aligned} \tag{34}$$

Thus, holonomies act as multiplicative operators, while the action of the smeared conjugated momentum depends upon a given edge γ of the graph crossing or not the associated surface S at the points P_k. For each intersection P_k the path γ is separated in two pieces $\gamma_1^{P_k}$ and $\gamma_2^{P_k}$ such that $\gamma = \gamma_1^{P_k} \circ \gamma_2^{P_k}$. The action at that intersection insert the generator σ_i between the holonomies corresponding to each of these paths. This is shown schematically in Fig. 3. To close this summary we mention the area operator $\hat{A}$ which is defined in the following way

$$\hat{A} = \lim_{n\to\infty} \sum_n \sqrt{\hat{E}_i(S_n)\hat{E}_i(S_n)} \ , \tag{35}$$

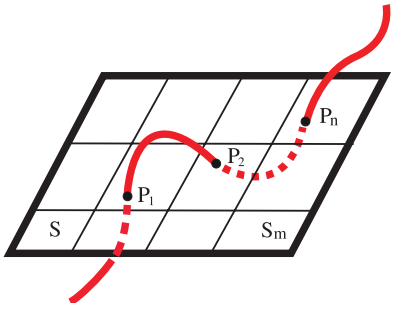


Fig. 3. Conditions for the action of the smeared momentum operator

by partitioning the associated surface S. A similar construction is made for the volume operator $\hat{V}$. These are well defined partial observables (hermitian operators in the $SU(2)$ gauge invariant kinematical space) with discrete eigenvalues and eigenfunctions given by the spin network states. An important property to be used in the following is that only the nodes contribute to the action of the volume operator upon states defined in a graph.

2.3 Elements of Thiemann Regularization

Thiemann has proposed a general regularization scheme that produces a consistent mathematical definition for the operators entering in the description of loop quantum gravity [61,62]. Such regularized operators act upon states which are functions of generalized connections defined over graphs. The regularization procedure is based upon a triangulation of space which is adapted to each graph. This means that the space surrounding any vertex of Γ is filled with tetrahedra Δ having only one vertex in common with the graph (called the basepoint $\mathbf{V}(\Delta)$) plus segments $\mathbf{U}_I(\Delta), I = 1,2,3$, starting at $\mathbf{V}$ and directed along the edges of the graph. For a vertex of valence greater than three one must consider all possible combination of three edges to build the tetrahedra. In the regions not including the vertices of Γ the choice of tetrahedra is arbitrary and the results are independent of it. The open path along the segment $\mathbf{U}_I$ is denoted by $\gamma_{\mathbf{V},\mathbf{U}_I}$. The arc connecting the end points of $\mathbf{U}_I(\Delta)$ and $\mathbf{U}_J(\Delta)$ is denoted by $a_{IJ}(\Delta)$ and the loop $\alpha_{\mathbf{V},\mathbf{U}_I,\mathbf{U}_J} := \gamma_{\mathbf{V},\mathbf{U}_I} \circ a_{IJ} \circ \gamma^{-1}_{\mathbf{V},\mathbf{U}_J} := \alpha_{IJ}(\Delta)$ can be formed. This construction is illustrated in the basic tetrahedron depicted in Fig. 4.

A first step in the implementation of the regularization method is to express each connection and field strength in terms of an adequate holonomy. For example, using the path $\gamma_{\mathbf{V},\mathbf{U}_1}$, of coordinate length ϵ, we can incorporate a connection $A_a(\mathbf{x})$ into the corresponding holonomy via the expansion

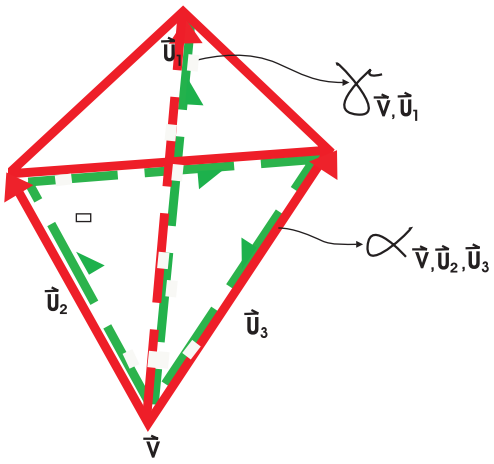


Fig. 4. Basic tetrahedron and paths in Thiemann regularization

$$U(\gamma_{\mathbf{V},\mathbf{U_1}}, A) = 1 + \epsilon U_1^a A_a(x) + O(\epsilon^2) \ . \tag{36}$$

An analogous calculation for the closed triangular path $\alpha_{\mathbf{V},\mathbf{U}_2,\mathbf{U}_3}$ leads to the following expression incorporating the curvature

$$U(\alpha_{\mathbf{V},\mathbf{U}_2,\mathbf{U}_3}, A) = 1 + \frac{1}{2}\epsilon^2 U_2^a \, U_3^b \ F_{ab} + O(\epsilon^3) \ . \tag{37}$$

A second ingredient of the method is the basic identity [61]

$$\frac{E_i^a(\mathbf{x}) E_j^b(\mathbf{x}) \epsilon^{ijk}}{\sqrt{det(E(\mathbf{x}))}} = \frac{2}{\kappa} \epsilon^{abc} \{A_c^k, V(R)\}_{PB}, \quad V = \int_R (d^3x) \sqrt{det(E^{ai})} \ , \tag{38}$$

where R is a region such $\mathbf{x} \in R$. In this way the volume V, which will be promoted to the operator level once the quantization is performed, is introduced in the regularization.

To fix ideas let us look at the first piece of the smeared Hamiltonian constraint

$$\mathcal{H}_1 = \int (d^3x) \, N(x) \, \frac{F_{ab}^k \epsilon^{ij}{}_k \, E_i^a E_j^b}{\sqrt{det(E)}} = \int (d^3x) N(x) \, \epsilon^{abc} F_{ab}^i \, \{A_{ci}, V\}) \ , \tag{39}$$

where we have used (38) in the second step. Now, substituting the expressions (36) and (37) for A_a, $F_{ab} \in SU(2)$, respectively, and replacing the integral with a sum, we can write

$$\mathcal{H}_1 = \lim_{\epsilon \to 0} \sum_m \epsilon^3 \, N_m \, \epsilon^{IJK} \frac{1}{\epsilon^3} Tr \Big(\ \big(U(\alpha_{\mathbf{V}_m,\mathbf{U}_I,\mathbf{U}_J}, A) - U^{-1}(\alpha_{\mathbf{V}_m,\mathbf{U}_I,\mathbf{U}_J}, A) \big) \\ \times U^{-1}(\gamma_{\mathbf{V}_m,\mathbf{U}_K}, A) \, \{U(\gamma_{\mathbf{V}_m,\mathbf{U}_K}, A), V(R_m)\} \Big) \ . \tag{40}$$

The crucial fact is that the divergent contribution arising from $1/\sqrt{det(E)}$, which goes like $1/\epsilon^3$, is cancelled by the correspondent factors contained in expansion of the holonomies. In this way (40) leads to a well defined regulated expression. Furthermore, when the correspondent quantities are quantized, the volume operator $\hat{V}$ is naturally introduced. This has the important consequence that the action of such operators upon wave functions defined on a graph get contributions only from the nodes of the graph.

Another illustrative example of this regularization is the magnetic sector of the electromagnetic Hamiltonian

$$H^B_{\text{Maxwell}} = \int_\Sigma d^3x \, \frac{q_{ab}}{2Q^2 \sqrt{\det q}} \underline{B}^a \underline{B}^b \ , \tag{41}$$

which is translated into the operator expression

$$\hat{H}^B = \frac{1}{2 \, \ell_P^4 \, Q^2} \sum_{\mathbf{V} \in V(\Gamma)} \ \sum_{\mathbf{V}(\Delta)=\mathbf{V}(\Delta')=\mathbf{V}} \epsilon^{JKL} \, \epsilon^{MNP} \\ \times \hat{w}_{iL\Delta} \ \hat{w}_{iP\Delta'} \ \Big(\underline{h}_{\alpha_{JK}(\Delta)} - 1 \Big) \ \Big(\underline{h}_{\alpha_{MN}(\Delta')} - 1 \Big) \ , \tag{42}$$

where

$$\hat{w}_{kI\Delta} = Tr\left(\tau_k h_{\mathbf{U}_I(\Delta)}\left[h^{-1}_{\mathbf{U}_I(\Delta)}, \sqrt{\hat{V}_v}\right]\right) . \tag{43}$$

Here we have simplified the notation by introducing

$$h_{\mathbf{U}_I(\Delta)} = U(\gamma_{\mathbf{V}_\Delta,\mathbf{U}_I}, A), \qquad \underline{h}_{\alpha_{JK}(\Delta)} = U(\alpha_{\mathbf{V}_\Delta,\mathbf{U}_J,\mathbf{U}_K}, \underline{A}) . \tag{44}$$

The underlined quantities (holonomy and connection) refer to the electromagnetic sector, and are to be distinguished from the gravitational ones.

In the following we will separately consider each matter contribution to the Hamiltonian constraint as the corresponding Hamiltonian for the respective gravitationally coupled sector.

3 A Kinematical Estimation of the Semiclassical Limit

The construction of states that live in the physical space of LQG and which approximate a given geometry at large distances, while retaining their granular structure at Planck length scales is still an open problem. This is usually referred to as the semiclassical limit of LQG and it is actively under investigation [63].

3.1 Heuristic Characterization of the States

In order to make some preliminary steps towards the study of the consequences of space granularity at large scales in the flat space limit, we take here an heuristical point of view, starting from the exact and well understood operator version of LQG. The approximation we introduce consists in defining their action upon the semiclassical states through some plausible requirements, without having an explicit construction for such states. We think of the semiclassical configuration corresponding to a particular matter or gauge field, described by operators collectively denoted by $\hat{F}(\mathbf{x})$, plus gravity as given by an ensemble of graphs Γ, each occurring with probability $P(\Gamma)$. To each of such graphs we associate a wave function $|\Gamma, \mathcal{L}, F\rangle := |\Gamma, S\rangle$ which we assume to be peaked with respect to the classical matter field configuration $F(\mathbf{x})$, together with a flat gravitational metric and a zero value for the gravitational connection at large distances. Not surprisingly, the semiclassical approximation gives us simultaneous average information regarding a field together with its canonically conjugated momentum. In other words, the contribution of the gravitational and matter operators inside the expectation value is estimated as [31–33]

$$\begin{aligned}
&\langle\Gamma, S|\ldots\hat{q}_{ab}\ldots|\Gamma, S\rangle \approx \delta_{ab} + O\left(\ell_P/\mathcal{L}\right),\\
&\langle\Gamma, S|\ldots\hat{A}_{ia}\ldots|\Gamma, S\rangle \approx \left(0 + (1/\mathcal{L})\left(\ell_P/\mathcal{L}\right)^{\Upsilon}\right)_{ia},\\
&\langle\Gamma, S|\ldots\underline{\hat{F}}_{ab}\ldots|\Gamma, S\rangle \approx \mu\underline{F}_{ab}
\end{aligned} \tag{45}$$

and similarly for any product of operators inside the expectation value. As an illustration, the gauge field is taken to be the electromagnetic field in the following

but an analogous estimation will hold for any other gauge or matter fields. The parameter $\Upsilon \geq 0$ is a real number and $\underline{F}_{ab}$ denotes the classical electromagnetic field strength. We further associate the effective hamiltonian

$$H_\Gamma = \langle \Gamma, S \,|\hat{H}_\Gamma|\, \Gamma, S\rangle \tag{46}$$

to each graph.

The coarse graining scale $\mathcal{L} \gg \ell_P$ of the wave function is such that a continuous flat metric approximation is appropriate for distances D much larger that $\mathcal{L}$, while the granular structure of space becomes relevant when probing distances D smaller that $\mathcal{L}$. That is to say, $\mathcal{L}$ is not a scale of quantum gravity but rather it is a scale that separates the continuum description of space from its discrete characterization that is fully manifest at the Planck length ℓ_P, which signals the quantum behavior. Summarizing, we expect the following behavior

$$\begin{aligned} &D \gg \mathcal{L} \gg \ell_P : \text{ continuous flat classical geometry,} \\ &\ell_P \sim D \ll \mathcal{L} \ : \text{manifest discrete structure of space ,} \end{aligned}$$

for the probe scale D. The coarse graining scale $\mathcal{L}$ is not provided by our approximation and has to be estimated in each particular case. We will explore some alternatives in Subsect. 3.4.

As indicated in Fig. 5, we think of space as constructed by adding boxes of size $\mathcal{L}^3$, which center represents a given point $\mathbf{x}$ in the continuum limit and which contain a large number of vertices of the adapted graphs. The matter field $F(\mathbf{x})$, characterized by a de Broglie wave length λ, is considered as a slowly varying function within each box ($\lambda > \mathcal{L}$) and contributes with its classical value at the center of the box, when taking expectation values. The requirement $\lambda > \mathcal{L}$ guarantees that we can describe the flat space dynamics in terms of an effective field theory, using the standard differential calculus of the continuum. On the other hand, gravitational variables are rapidly varying inside each box.

The total effective Hamiltonian is defined as an average over the graphs Γ

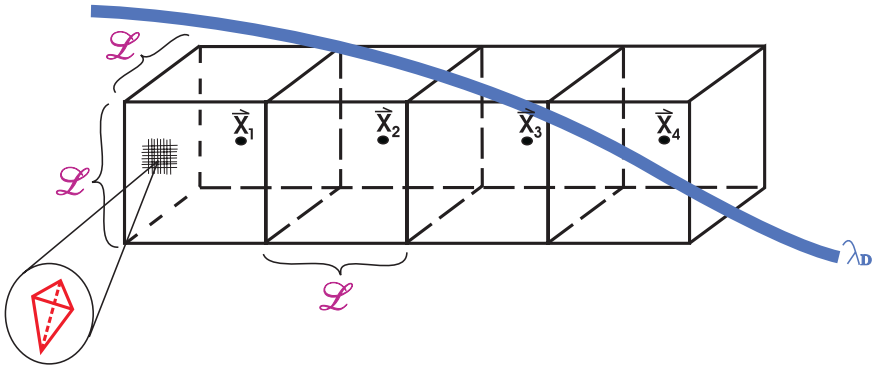


Fig. 5. Different scales for the model of space considered in the text

$$\mathrm{H} = \sum_{\Gamma} P(\Gamma)\, \mathrm{H}_{\Gamma} \ . \tag{47}$$

3.2 The Calculation

We summarize now the method of calculation [31–33]. For each graph Γ the effective Hamiltonian is defined by (46). For a given vertex $\mathbf{V}$ inside the expectation value, one expands each holonomy or flux operator in powers of the segments $U^a_I(\Delta))$ of the attached tetrahedra Δ, plus derivatives of the gauge or matter fields operators. Schematically, in the case of (42) this produces

$$H^B_\Gamma = \sum_{\mathbf{V}\in \mathbf{V}(\Gamma)} \sum_{\mathbf{V}(\Delta)=\mathbf{V}} \langle \Gamma, S| \left(\partial^{a_1}\partial^{a_2} \dots \partial^{a_k} \dots \underline{\hat{F}}_{p_1 q_1}(\mathbf{V}) \right) \dots$$
$$\times \left(\dots \partial^{a_{m-1}}\partial^{a_m} \underline{\hat{F}}_{pq}(\mathbf{V}) \right) \hat{T}^{p_1\, q_1 \dots p\, q}_{a_1 \dots a_m}(\mathbf{V}, \mathbf{U}_I(\Delta), \hat{A}_{ia}, \hat{V}) \ |\Gamma, S\rangle \ . \tag{48}$$

Here $\hat{T}$ contains gravitational operators (connection and volume operators, for example) together with contributions depending on the segments of the adapted triangulation in the particular graph. Next, according to Fig. 5, space is considered to be divided into boxes, each centered at a fixed point $\mathbf{x}$ and with volume $\mathcal{L}^3 \approx d^3\, x$. The choice of boxes is the same for all the graphs considered. Each box contains a large number of vertices of the semiclassical state ($\mathcal{L} \gg \ell_P$), but it is considered as infinitesimal in the scale where the space can be regarded as continuous. The sum over the vertices in (48) is subsequently split as a sum over the vertices in each box, plus a sum over the boxes. Also, one assumes that the gauge or matter operators are slowly varying within a box ($\mathcal{L} \ll \lambda$, with λ been the correspondent particle wavelength), in such a way that for all the vertices inside a given box one can write $\langle \Gamma, S| \dots \underline{\hat{F}}_{ab}(\mathbf{V}) \dots |\Gamma, S\rangle = \mu \underline{F}_{ab}(\mathbf{x})$, for example. Here $\underline{F}_{ab}$ is the classical electromagnetic field at the center of the box and μ is a dimensionless constant which is determined in such a way that the standard classical result in the zeroth order approximation is recovered. Applying the procedure just described to (48) leads to

$$H^B_\Gamma = \sum_{\mathrm{Box}} \left(\partial^{a_1}\partial^{a_2} \dots \partial^{a_k} \dots \underline{F}_{p_1 q_1}(\mathbf{x}) \right) \dots \left(\dots \partial^{a_{m-1}}\partial^{a_m} \underline{F}_{pq}(\mathbf{x}) \right)$$
$$\sum_{\mathbf{V}\in \mathrm{Box}} {\ell_P}^3 \sum_{\mathbf{V}(\Delta)=\mathbf{V}} \mu^{n+1} \langle \Gamma, S| \frac{1}{\ell_P^3} \hat{T}_{a_1 \dots a_m}{}^{pq \dots p_1 q_1}(\mathbf{V}, \mathbf{U}(\Delta), A_{ia}) |\Gamma, S\rangle \ , \tag{49}$$

where $n+1$ is the total number of factors $F_{rs}(\mathbf{x})$, each of which can contain some derivatives. The expectation value of the gravitational contribution is supposed to be a rapidly varying function inside each box. Next we consider the total effective Hamiltonian (47), which is defined as an average over the graphs Γ, i.e. over the adapted triangulations. This effectively amounts to average the

remaining expectation values in each box of the sum (49). We call this box-average $t_{a_1\ldots a_m}{}^{pq\ldots p_1 q_1}(\mathbf{x})$ and define it by

$$\sum_{\Gamma} P(\Gamma) \sum_{\mathbf{V}\in \mathrm{Box}} {\ell_P}^3 \sum_{\mathbf{V}(\Delta)=\mathbf{V}} \langle \Gamma, S| \frac{1}{\ell_P^3} \hat{T}_{a_1\ldots a_m}{}^{pq\ldots p_1\, q_1} |\Gamma, S\rangle$$
$$= d^3x\, t_{a_1\ldots a_m}{}^{pq\ldots p_1 q_1}(\mathbf{x}) \; . \tag{50}$$

We estimate $t_{a_1\ldots a_m}{}^{pq\ldots p_1 q_1}(\mathbf{x})$ by demanding it to be constructed from the flat space tensors δ_{ab} and ϵ_{abc}, together with the corresponding ones τ_i and δ_i^a, in the case of spinors. In this way we are imposing isotropy and rotational invariance on our final Hamiltonian in the frame selected by the choice of the semiclassical states. This is somewhat analogous to the spontaneously symmetry breaking scheme, where the choice of the vacuum selects a particular frame for the symmetry breaking vacuum expectation values. Also, the scalings given in (45) together with the additional assumptions

$$\langle \Gamma, S| \ldots \hat{V} \ldots |\Gamma, S\rangle \longrightarrow \ell_P^3, \qquad U_I^a \longrightarrow \ell_P \; , \tag{51}$$

are used in the above estimation. After replacing the summation over boxes by the integral over space, the resulting Hamiltonian has the final form

$$H^B = \int d^3x \, \Big(\partial^{a_1}\partial^{a_2}\ldots\partial^{a_k}\ldots \underline{F}_{p_1 q_1}(\mathbf{x})\Big)\ldots \Big(\ldots \partial^{a_{m-1}}\partial^{a_m} \underline{F}_{pq}(\mathbf{x})\Big)$$
$$\times t_{a_1\ldots a_m}{}^{pq\ldots p_1\, q_1}(\mathbf{x}) \; . \tag{52}$$

Finally we exhibit some of the required expansions. The basic electromagnetic holonomy around the path α_{JK}

$$\underline{h}_{\alpha_{JK}(\Delta)} = e^{-i\int_{\alpha_{JK}(\Delta)} dt\, \dot{U}^a(t) \underline{\hat{A}}_a(s(t))} = e^{-i\Phi^B(F_{JK})} \; , \tag{53}$$

can be written in terms of the magnetic flux Φ^B through the face bounded by such closed path. The expansion of such flux with respect to the magnetic field at the vertex is

$$\Phi^B(F_{IJ}) = \Big(1 \; + \frac{1}{3}(U_I^c + U_J^c)\, \partial_c \; + \frac{1}{12}(U_I^c\, U_I^d + U_I^c\, U_J^d + U_J^c\, U_J^d)$$
$$\times \; \partial_c\, \partial_d + \; \ldots \Big) \; \frac{1}{2} U_I^a U_J^b \epsilon_{abc} B^c(\mathbf{V}) \; ,$$

to second order in the derivatives.

The expansion of the gravitational operator (43) is

$$\hat{w}_{i\,L\Delta} = U_L^a w_{ia} + U_L^a U_L^b w_{iab} + U_L^a U_L^b U_L^c w_{iabc} + \mathcal{O}(s^4 w) \; , \tag{54}$$

to third order in the vectors U_L^a. The remaining operators in the expansion include the connection together with the volume operators, in the form

$$w_{ia} = \frac{1}{2}[A_{ia}, \sqrt{V}], \quad w_{iab} = \frac{1}{8}\epsilon_{ijk}[A_{ja}, [A_{kb}, \sqrt{V}]],$$
$$w_{iabc} = -\frac{1}{48}[A_{ja}, [A_{jb}, [A_{ic}, \sqrt{V}]]] \; . \tag{55}$$

3.3 The Results

The application of the method described in the previous section leads to the following effective Hamiltonians. In the electromagnetic case we obtain [32]

$$
\begin{aligned}
H^{EM} = \frac{1}{Q^2}\int d^3\mathbf{x}\Bigg[& \left(1+\theta_7\left(\frac{\ell_P}{\mathcal{L}}\right)^{2+2\Upsilon}\right)\frac{1}{2}\left(\underline{\mathbf{B}}^2+\underline{\mathbf{E}}^2\right) \\
& +\theta_8\ell_P\left(\underline{\mathbf{B}}\cdot(\boldsymbol{\nabla}\times\underline{\mathbf{B}})+\underline{\mathbf{E}}\cdot(\boldsymbol{\nabla}\times\underline{\mathbf{E}})\right) \\
& +\theta_3\,\ell_P^2\left(\underline{B}^a\,\nabla^2\underline{B}_a+\underline{E}^a\,\nabla^2\underline{E}_a\right)+\theta_2\,\ell_P^2\,\underline{E}^a\partial_a\partial_b\underline{E}^b \\
& +\theta_4\,\ell_P^2\mathcal{L}^2\left(\frac{\ell_P}{\mathcal{L}}\right)^{2\Upsilon}\left(\underline{\mathbf{B}}^2\right)^2+\ldots\Bigg]\,. \qquad (56)
\end{aligned}
$$

The above result has been extended to the Yang-Mills case in [64]. The case where only θ_8 in non zero corresponds to the Gambini-Pullin effective Hamiltonian [30].

In the case of Majorana fermions with mass m we have [31, 33]

$$
\begin{aligned}
H^F = \int d^3x\Bigg[& i\,\pi(\mathbf{x})\,\left(1+\kappa_1\left(\frac{\ell_P}{\mathcal{L}}\right)^{1+\Upsilon}+\kappa_2\left(\frac{\ell_P}{\mathcal{L}}\right)^{2+2\Upsilon}\right. \\
& \left.+\frac{\kappa_3}{2}\,\ell_P^2\,\nabla^2\right)\tau^d\partial_d\,\xi(\mathbf{x}) \\
& +\frac{i}{4}\frac{1}{\mathcal{L}}\,\pi(\mathbf{x})\left(\kappa_4\left(\frac{\ell_P}{\mathcal{L}}\right)^{\Upsilon}+\kappa_5\left(\frac{\ell_P}{\mathcal{L}}\right)^{1+2\Upsilon}+\cdots+\frac{\kappa_7}{2}\,\ell_P^2\nabla^2\right)\xi(\mathbf{x}) \\
& +\frac{m}{2}\xi^T(\mathbf{x})\,(i\sigma^2)\left(1+\kappa_8\left(\frac{\ell_P}{\mathcal{L}}\right)^{1+\Upsilon}+\left(\kappa_9\ell_P+\ldots\right)\tau^a\partial_a\right)\xi(\mathbf{x})+c.c.\Bigg]\,, \qquad (57)
\end{aligned}
$$

where $\pi(\mathbf{x}) = i\xi^*(\mathbf{x})$ is the canonically conjugated momentum of the two-component spin 1/2 field $\xi(\mathbf{x})$. The notation is $\tau^a = -(i/2)\sigma^a$, where σ^a are the standard Pauli matrices.

Some comments regarding the procedure are now in order. (i) The dimensionless numerical coefficients θ_A, κ_A appearing in the above effective Hamiltonians remain arbitrary in the procedure, but are independent of any parameter of the model. To predict them will require an exact knowledge of the states used to calculate the corresponding average values, together with the action of the basic operators upon them. In this respect we have obtained a systematic parameterization of the possible modified Hamiltonians in terms of higher derivative operators, where the dependence upon the scales of the problem $(\ell_P, \mathcal{L}, \Upsilon)$ has been explicitly determined. In particular, these coefficients could turn out to be zero, leading to no dynamical corrections. (ii) A main drawback of the method is that it does not incorporate properly the dynamics via the Hamiltonian constraint of quantum gravity. An improved semiclassical approximation taking care of the many issues assumed or left over in our approach is certainly needed. (iii)

There is also the question of interpreting the results (56) and (57) in relation to their transformation properties under standard active (particle)global Lorentz transformations. There are various possibilities. These Hamiltonians have been calculated in a specific reference frame where isotropy under rotations is maintained. This could signal the presence of an absolute reference frame, a reborn version of the ether. This hypothesis will be further explored in the following and has received a lot of attention in the framework of different models in the literature. On the other hand one could argue that this specific frame has been selected by the choice of the semiclassical states that describe the flat space limit, in analogous way to the spontaneously symmetry breaking procedure. The transformation properties under this assumption have not been studied. The third alternative is that only a complete calculation of the corrections to flat space dynamics will tell us which is the proper generalization, if any, of the global Lorentz transformations, thus yielding a deformed or extended version for preserving the equivalence of inertial frames and leading to a modified relativity principle. (iv) Finally we must say that the above results have provided some motivation for phenomenological theories which have explored the consequences of such modifications and have shown that either experimental or astrophysical observations, even with the actual level of sensitivity, can set rather stringent bounds upon the correction parameters. In other words, Planck scale sensitivity has been already attained in a rather broad set of observations. Conversely, this would imply severe constraints upon the fundamental theory once the semiclassical limit is correctly performed.

3.4 The Parameters $\mathcal{L}$ and Υ

In order to produce numerical estimations for some of the effects arising from the previously obtained modifications to flat space dynamics, we must further fix the value of the scales $\mathcal{L}$ and Υ.

Let us recall that $\mathcal{L}$ is a scale indicating the onset of the distance from where the non perturbative states in the loop representation can be approximated by the classical flat metric. The propagating particle is characterized by energies which probe distances of the order of the De Broglie wave length λ. As mentioned previously, just to be consistent with a description in terms of classical continuous equations it is necessary to require that $\mathcal{L} < \lambda$. Two distinguished cases for choosing $\mathcal{L}$ arise: (i) the mobile scale, where we take the marginal choice $\mathcal{L} = \lambda$ in each situation and (ii) the universal scale, which has been introduced in the discussion of the GZK anomaly [50]. The consideration of the different reactions involved produces the preferred range $4.6\times10^{-8}\,\mathrm{GeV}^{-1} \geq \mathcal{L} \geq 8.3\times10^{-9}\,\mathrm{GeV}^{-1}$. A recent study of the gravitational Cerenkov effect together with neutrino oscillations [65] yields a universal scale evaluation which is consistent with the former.

Ranges for Υ have been estimated considering the observation that atmospheric neutrino oscillations at average energies of the order $10^{-2} - 10^{2}$ GeV are dominated by the corresponding mass differences via the oscillation length L_m [66]. This means that additional contributions to the oscillation length, in

Table 2. Ranges for the parameters $\mathcal{L}$ and Υ

	$\mathcal{L}\,[\mathrm{GeV}^{-1}]$	Υ
Mobile scale	$\mathcal{L} = \lambda$	$\Upsilon > 0.15$
Universal scale	$8.3 \times 10^{-9} < \mathcal{L} < 4.6 \times 10^{-8}$	$\Upsilon > 1.2$

particular the quantum gravity correction L_{QG}, should satisfy $L_{QG} > L_m$. This has been used to set lower bounds upon Υ. Within the proposed two methods for estimating the scale $\mathcal{L}$ of the process we obtain: (i) $\Upsilon > 0.15$ when $\mathcal{L}$ is considered as a mobile scale and (ii) $\Upsilon > 1.2$ when the scale $\mathcal{L}$ takes the universal value $\mathcal{L} \approx 10^{-8}\,\mathrm{GeV}^{-1}$ [33]. These results are summarized in Table 2.

4 Phenomenological Aspects

In this section we discuss four phenomenological applications of models presenting Planck scale dynamical corrections. In Subsect. 4.1 we summarize the derivation of stringent bounds upon combinations of some parameters appearing in the effective fermionic Hamiltonian (57), using existing data from clock-comparison experiments. Subsection 4.2 is devoted to the calculation of radiative corrections arising in LIV effects with preferred frames associated to space granularity. In order to make the calculation simple we consider a Yukawa model for the interactions. Subsection 4.3 contains the discussion of synchrotron radiation in Planck scale modified electrodynamics, emphasizing the model of [14] which is phenomenologically formulated as a theory parameterizing LIV with dimension five operators. Finally, a unified description of radiation in Planck scale modified electrodynamics incorporating different models is presented in Subsect. 4.4.

4.1 Bounds on the Fermionic Sector Parameters from Clock-Comparison experiments

The Hamiltonian (57) for two components fermions was obtained under the assumption of flat space isotropy and it was assumed to account for the dynamics in a preferred reference frame, identified as the one in which the Cosmic Microwave Background looks isotropic. The earth velocity $\mathbf{w}$ with respect to that frame has already been determined to be $w/c \approx 1.23 \times 10^{-3}$ by COBE [67]. Thus, in the earth reference frame one expects the appearance of signals indicating minute violations of space isotropy encoded in the diurnal variation of the $\mathbf{w}$-dependent terms appearing in the transformed Hamiltonian or Lagrangian [68]. On the other hand, many high precision experimental test of such variations, using atomic and nuclear systems for example, have been already reported in the literature [10, 69–71] and already analyzed in terms of the SME. Amazingly enough such precision is already adequate to set very stringent bounds on some of the parameters arising from the quantum gravity corrections.

We have considered the case of non-relativistic Dirac particles and obtained corrections which involve the coupling of the spin to the CMB velocity, together with a quadrupolar anisotropy of the inertial mass [68]. The calculation was made with the choices $\Upsilon = 0$ and $\mathcal{L} = 1/M$, where M is the rest mass of the fermion. Here it is important to emphasize again that $\mathcal{L} \gg \ell_P$ is not the scale of quantum gravity but rather the lower limit of distances from which a description of space as a continuum is already valid. In this sense we are taking the upper limit settled by the de Broglie wave length $\lambda_D = 1/M$ of the nucleon, which allows us to make sense of the corresponding Dirac equation. As we will show later, a lower limit for $\mathcal{L}$ will only make the bounds upon the LIV parameters more stringent.

Keeping only terms linear in ℓ_P, the equations of motion for a Majorana fermion of mass m described by the two-component spinor ξ, arising from the Hamiltonian (57) are given by

$$\begin{aligned} &[i\partial/\partial t - iA\boldsymbol{\sigma}\cdot\boldsymbol{\nabla} + K/2]\,\xi - m(\alpha - i\beta\boldsymbol{\sigma}\cdot\boldsymbol{\nabla})\chi = 0\,, \\ &[i\partial/\partial t + iA\boldsymbol{\sigma}\cdot\boldsymbol{\nabla} - K/2]\,\chi - m(\alpha - i\beta\boldsymbol{\sigma}\cdot\boldsymbol{\nabla})\xi = 0\,, \end{aligned} \tag{58}$$

where

$$\begin{aligned} &A = (1 + \Theta_1 m\ell_p), \quad \alpha = (1 + \Theta_3 m\ell_P)\,, \\ &K = m\Theta_4 m\ell_P, \quad \beta = \Theta_2\ell_P\,. \end{aligned} \tag{59}$$

The notation in (58) is $\chi = -i\sigma_2\,\xi^*$ and it is a direct matter to verify the consistency between them. These equations can be readily extended to the Dirac case by considering χ and ξ as independent spinors unified in $\Psi^T = (\xi^T,\, \chi^T)$, with the result

$$\left(i\gamma^\mu\partial_\mu + \Theta_1 m\ell_P\; i\gamma\cdot\nabla - \frac{K}{2}\gamma_5\gamma^0 - m\,(\alpha - i\Theta_2\ell_P\,\Sigma\cdot\nabla)\right)\Psi = 0\,. \tag{60}$$

Here we have used the representation in which γ_5 is diagonal and the spin operator is $\Sigma^k = (i/2)\epsilon_{klm}\gamma^l\gamma^m$, with standard particle physicist conventions. The normalization has been chosen so that in the limit $(m\ell_P) \to 0$ we recover the standard massive Dirac equation. The term $m\,(1 + \Theta_3\, m\ell_P)$ can be interpreted as a renormalization of the mass whose physical value is taken to be $M = m\,(1 + \Theta_3\, m\ell_P)$. After these modifications we can write an effective Lagrangian describing the time evolution as seen in the CMB frame. In order to obtain the dynamics in the laboratory frame we implement an observer (passive) Lorentz transformation in the former Lagrangian and rewrite it in a covariant looking form by introducing explicitly the CMB frame four velocity $W^\mu = \gamma(1,\, \mathbf{w}/c)$ arising from the boosted rest frame velocity $(1, \mathbf{0})$. The result is

$$\begin{aligned} L_D = {} & \frac{1}{2}i\bar\Psi\gamma^\mu\partial_\mu\Psi - \frac{1}{2}M\bar\Psi\,\Psi + \frac{1}{2}i(\Theta_1 M\ell_P)\bar\Psi\gamma_\mu\,(g^{\mu\nu} - W^\mu W^\nu)\,\partial_\nu\Psi \\ & + \frac{1}{4}(\Theta_2 M\ell_P)\bar\Psi\epsilon_{\mu\nu\alpha\beta}W^\mu\gamma^\nu\gamma^\alpha\partial^\beta\Psi - \frac{1}{4}(\Theta_4 M\ell_P)MW_\mu\bar\Psi\gamma_5\gamma^\mu\Psi + h.c.\,, \end{aligned} \tag{61}$$

where W^μ is an external non-dynamical field which is not to be varied in the corresponding action. It is interesting to remark that the above Lagrangian provides a specific realization of the general form considered in the SME. According to [72], the identifications are

$$a_\mu = H_{\mu\nu} = d_{\mu\nu} = e_\mu = f_\mu = 0, \qquad c_{\mu\nu} = \Theta_1 M\ell_P(g_{\mu\nu} - W_\mu W_\nu) \,,$$
$$g_{\alpha\beta\gamma} = -\Theta_2 M\ell_P W^\rho \epsilon_{\rho\alpha\beta\gamma}, \qquad b_\mu = \frac{1}{2}\Theta_4 M^2 \ell_P W_\mu \,. \tag{62}$$

The above reference provides also the non-relativistic limit of the Hamiltonian corresponding to the general SME modified fermion Lagrangian. In our case, up to first order in ℓ_P and up to order $(\mathbf{w})/c^2$, the identifications (62) yield

$$\begin{aligned}\tilde{H} = {} & Mc^2(1 + \Theta_1\, M\ell_P\, (\mathbf{w}/c)^2) + \Theta_1 M\ell_P \left[\frac{\mathbf{w}\cdot Q_P\cdot\mathbf{w}}{Mc^2}\right] \\ & + \left(1 + 2\,\Theta_1 M\ell_P \left(1 + \frac{5}{6}\,(\mathbf{w}/c)^2\right)\right)\left(\frac{p^2}{2M} + g\,\mu\,\mathbf{s}\cdot\mathbf{B}\right) \\ & + \left(\Theta_2 + \frac{1}{2}\Theta_4\right) M\ell_P \left[\left(2Mc^2 - \frac{2p^2}{3M}\right)\mathbf{s}\cdot\frac{\mathbf{w}}{c} + \frac{1}{M}\,\mathbf{s}\cdot Q_P\cdot\frac{\mathbf{w}}{c}\right] \,, \end{aligned} \tag{63}$$

where $s^i = \sigma^i/2$.

The above effective Hamiltonian has been employed in the description of the valence nucleon responsible for the transitions measured in clock-comparison experiments using pairs of nuclei like (^{21}Ne, ^{3}He) [70], and (^{129}Xe, ^{3}He) [71], for example. In (63) we have not written the terms linear in the momentum since they average to zero in the nuclear bound state situation. Here g is the standard gyromagnetic factor, and Q_P is the momentum quadrupole tensor with components $Q_{Pij} = p_i p_j - 1/3 p^2\delta_{ij}$. The terms in the second square bracket of the LHS of (63) represent a coupling of the spin to the velocity of the privileged reference frame. The first term inside the bracket has been measured with high accuracy and an upper bound for the coefficient has been found. The second term in the same bracket is a small anisotropy contribution and can be neglected. Thus we find the correction

$$\delta H_S = \left(\Theta_2 + \frac{1}{2}\Theta_4\right) M\ell_P(2Mc^2)\left[1 + O\left(\frac{p^2}{2M^2c^2}\right)\right]\mathbf{s}\cdot\frac{\mathbf{w}}{c} \,. \tag{64}$$

The first square bracket in the LHS of (63) represents an anisotropy of the inertial mass and has been bounded in Hughes-Drever like experiments. With the approximation $Q_P = -5/3 < p^2 > Q/R^2$ for the momentum quadrupole moment, with Q being the electric quadrupole moment and R the nuclear radius, we obtain

$$\delta H_Q = -\Theta_1 M\ell_P \frac{5}{3}\left\langle\frac{p^2}{2M}\right\rangle\left(\frac{Q}{R^2}\right)\left(\frac{w}{c}\right)^2 P_2(\cos\theta) \,, \tag{65}$$

for the quadrupole mass perturbation, where θ is the angle between the quantization axis and $\mathbf{w}$. Using $\langle p^2/2M\rangle \sim 40$ MeV for the energy of a nucleon in the

outer shell of a typical heavy nucleus, together with the experimental bounds of references [70, 71] we find [68]

$$\left|\Theta_2 + \frac{1}{2}\Theta_4\right| < 2 \times 10^{-9}, \qquad |\,\Theta_1\,| < 3 \times 10^{-5}\,. \tag{66}$$

From (64) and (65) we realize that only products of the type (LIV parameters) $\times\, 1/\mathcal{L} \approx$ (LIV parameters) $\times\, M$ are bounded by the experiments. In this way choosing a smaller $\mathcal{L}$, i.e. a larger M, will just decrease the upper bound for the LIV parameters. Thus the choice $\mathcal{L} = \lambda_D = 1/M$ produces the weakest bound in this analysis.

4.2 Radiative Corrections in Preferred Frames Modelling Space Granularity

So far corrections to the dynamics have arisen only in the non-interacting theory, through factors of the type $(E/M_P)^{\Upsilon}$, which are directly relevant at unaccessible energies $E \approx M_P$. A possibility of probing such high energies is through the inclusion of radiative corrections (particle's self energies, for example) because the internal momenta are integrated up to the maximum allowed in a given reference frame. The standard folklore with respect to any new physics entering at high scales (here Planck scale) is that it has negligible effects on the leading-order low-energy physics (here free particle corrections). Contrary to this belief, we show in this subsection that modelling space granularity via the introduction of a physical cutoff, which defines a preferred reference frame, leads to unsuppressed dimension four LIV contributions [74, 75]. Results consistent with these have been obtained in [14].

To this end we consider the calculation of one-loop radiative corrections in the Yukawa model

$$L = 1/2\,(\partial\phi)^2 - 1/2\,\mu^2\phi^2 + \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi + g\phi\bar{\psi}\psi + (\mathrm{LVT})\,. \tag{67}$$

Here LVT refers to the highly suppressed zeroth order Lorentz violating terms that take into account the previously discussed free particle dynamical modifications. We model the space granularity by introducing a physical cut-off Λ in such a way that the magnitude of the three-momentum in any loop is bounded by this quantity. The parameter Λ defines the onset of the scale at which the granularity of space becomes manifest. A convenient way of incorporating this requirement is to introduce the physical cutoff function

$$f(|\mathbf{k}|/\Lambda), \quad f(0) = 1, \quad f(\infty) = 0, \quad \Lambda \approx 1/\mathcal{L}\,, \tag{68}$$

which suppresses the internal momenta having $|\mathbf{k}| \geq \Lambda$. One can visualize this function as a smoothed theta function with the correspondent jump at $|\mathbf{k}| \approx \Lambda$. Our choice is that the cutoff function depends on spatial momentum, but not of energy, and that it is rotationally invariant.

The Fermion Self Energy [75]

Tests of Lorentz invariance typically deal with the relation between energy and 3-momentum of an isolated particle, which is given by the position of the pole of the particle's full propagator. We will calculate the effect of loop corrections on this relation. It is convenient, as usual, to write the full fermion propagator as

$$S(p) = \frac{i}{p^\mu \gamma_\mu - m - \Sigma_2(p)} \ , \tag{69}$$

where $\Sigma_2(p)$ is the standard fermion self energy. Here we are neglecting the zeroth order corrections to the dispersion relations for the free particles.

As a first manner to model the validity of a continuous description of space only up to some short distance $1/\Lambda \approx \mathcal{L}$, we choose to cutoff only the free scalar propagator, but not the free fermion propagator, which is sufficient to cutoff the UV divergence of the one-loop fermion self-energy while giving a maximally simple calculation. In this way, the one-loop approximation to the fermion self-energy is

$$\Sigma_2(p) = ig^2 \int \frac{\mathrm{d}^4 k}{(2\pi)^4} \, \frac{f(|\mathbf{k}|/\Lambda)}{k^2 + i\epsilon} \, \frac{\gamma_\rho (p+k)^\rho + m}{(p+k)^2 - m^2 + i\epsilon} \ , \tag{70}$$

where we have set $\mu = 0$ for simplicity. We wish to investigate the self-energy when the momentum p^μ and the mass m are much less than the parameter Λ. If the integral were convergent as the parameter is removed ($\Lambda \to \infty$), then the self-energy would equal its value at $\Lambda = \infty$ plus corrections of order m/Λ and p/Λ. However, it is easy to see that the integral is divergent in this limit. To analyze its behavior, we observe that differentiating it twice with respect to p gives a finite integral (when $\Lambda \to \infty$). So we extract the zeroth order and the linear terms in the momentum expansion of Σ_2 about $p = 0$

$$\Sigma_2(p) = \Sigma_2(p=0) + p^0 \partial_{p^0} \Sigma_2(p)|_{p=0} + p^i \partial_{p^i} \Sigma_2(p)|_{p=0} + \hat{\Sigma}(p) \ . \tag{71}$$

Now $\hat{\Sigma}(p)$ is $O(p^2)$ as $p \to 0$ and it is finite when the cutoff is removed. Therefore it only gives power-suppressed Lorentz violations. The unsuppressed Lorentz violations, if any, can only arise from the first three terms. We find that the terms $\Sigma_2(p=0)$, $\partial_{p^0} \Sigma_2(p)|_{p=0}$ and $\partial_{p^i} \Sigma_2(p)|_{p=0}$ are proportional to the Dirac matrices 1, γ_i, γ^0, exactly as in the Lorentz-invariant case; this follows from the discrete symmetries of our model, rather than full Lorentz invariance. Hence we can write

$$\begin{aligned} \Sigma_2(p) &= Am + p^0 \gamma_0 B + p^i \gamma_i C + \hat{\Sigma}(p, \Lambda = \infty) + O(p/\Lambda, m/\Lambda) \\ &= Am + p^\mu \gamma_\mu C + p^0 \gamma_0 (B - C) + \hat{\Sigma}(p, \Lambda = \infty) + O(p/\Lambda, m/\Lambda) \ , \end{aligned} \tag{72}$$

where A, B and C are numerical-valued functions of the parameters of the problem. The only leading-power Lorentz violation is caused by the difference between the B and C, which we calculate below and find it to be nonzero and unsuppressed.

In fact, the constant term is

$$\Sigma_2(0) = \frac{ig^2}{(2\pi)^4} \int \mathrm{d}^4k\, f(|\mathbf{k}|/\Lambda)\, \frac{\gamma_\rho k^\rho + m}{(k^2 - m^2 + i\epsilon)\,(k^2 + i\epsilon)}\ . \tag{73}$$

Since the physical cutoff depends only on $|\mathbf{k}|$, the modified theory is invariant under reversal of any component of k^μ and the term proportional to γ_ρ vanishes after integration. Hence

$$\begin{aligned} A &= \frac{ig^2}{(2\pi)^4} \int \mathrm{d}^4k\, \frac{f(|\mathbf{k}|/\Lambda)}{(k^2 - m^2 + i\epsilon)\,(k^2 + i\epsilon)}. \\ &= \frac{g^2}{4\pi^2} \int_0^\infty \mathrm{d}l\, f(l/\Lambda)\, \frac{l^2}{m^2} \left(\frac{1}{l} - \frac{1}{E_l}\right)\ , \end{aligned} \tag{74}$$

where $E_l = \sqrt{l^2 + m^2}$. In the second line, we have performed the integrals over k^0 and the angle of $\mathbf{k}$. The corresponding term in Σ_2 is proportional to the unit Dirac matrix, so that it is equivalent to a fermion mass term. Note that when $\Lambda \to \infty$ there is a logarithmic divergence, which can be removed by the usual mass renormalization. Since the term is proportional to the unit Dirac matrix, it does not violate Lorentz invariance.

The situation for the derivative terms is different. For the time-like derivative we have

$$\partial_{p^0} \Sigma_2(p)|_{p=0} = \frac{ig^2}{(2\pi)^4} \int \mathrm{d}^4k\, \frac{f(|\mathbf{k}|/\Lambda)}{k^2 + i\epsilon} \left[\frac{\gamma_0}{k^2 - m^2 + i\epsilon} - \frac{2k^0(\gamma_\rho k^\rho + m)}{(k^2 - m^2 + i\epsilon)^2}\right]\ . \tag{75}$$

Invariance under reversal of components removes all but the term proportional to γ_0, so that

$$\begin{aligned} B &= \frac{ig^2}{(2\pi)^4} \int \mathrm{d}^4k\, \frac{f(|\mathbf{k}|/\Lambda)}{k^2 + i\epsilon} \left[\frac{1}{k^2 - m^2 + i\epsilon} - \frac{2k^{0\,2}}{(k^2 - m^2 + i\epsilon)^2}\right] \\ &= \frac{g^2}{4\pi^2} \int_0^\infty \mathrm{d}l\, f(l/\Lambda) \left[\frac{2l^2(E_l - l)}{m^4} - \frac{l}{m^2}\right]\ . \end{aligned} \tag{76}$$

Similarly, from the space-like derivative we get

$$\begin{aligned} C &= \frac{ig^2}{(2\pi)^4} \int \mathrm{d}^4k\, \frac{f(|\mathbf{k}|/\Lambda)}{k^2 + i\epsilon} \left[\frac{1}{k^2 - m^2 + i\epsilon} + \frac{2\mathbf{k}^2/3}{(k^2 - m^2 + i\epsilon)^2}\right] \\ &= \frac{g^2}{4\pi^2} \int_0^\infty \mathrm{d}l\, f(l/\Lambda) \left[\frac{2l^3}{3m^4} - \frac{l}{m^2} + \frac{l^2}{m^2 E_l} - \frac{2l^4}{3m^4 E_l} - \frac{l^4}{3m^2 E_l^3}\right]\ . \end{aligned} \tag{77}$$

As expected, both B and C have logarithmic divergences as $\Lambda \to \infty$.

We are interested in the difference, which gives the low-energy violation of Lorentz invariance

$$B - C = -\frac{2ig^2}{(2\pi)^4} \int \mathrm{d}^4k \, \frac{f(|\mathbf{k}|/\Lambda)}{k^2 + i\epsilon} \, \frac{\frac{1}{3}\mathbf{k}^2 + {k^0}^2}{(k^2 - m^2 + i\epsilon)^2} \tag{78}$$

$$= \frac{g^2}{12\pi^2} \int_0^\infty \mathrm{d}l \, f(l/\Lambda) \left[\frac{8E_l^3}{m^4} - \frac{8l^3}{m^4} - \frac{12E_l}{m^2} + \frac{3}{E_l} + \frac{m^2}{E_l^3} \right] . \tag{79}$$

Suppose the integral (78) were convergent in the continuum limit. Then it would be zero in this limit, as can be seen by Wick rotation of an integral of $\frac{1}{3}\mathbf{k}^2 + {k^0}^2$ times an arbitrary Lorentz-invariant function $F(k^2)$. This argument depends on the integral being absolutely convergent, i.e., on the integral of the absolute value of the integrand being finite. In contrast, the corresponding integral for $B - C$ is logarithmically divergent. So the leading term for the integral, obtained when $\Lambda \to \infty$, depends on how this cutoff is removed. In our case this is dictated by the physical modelling of the space granularity via the function (68), as expressed in (79). This choice implies that the factor in square brackets behaves like m^2/l^3 when $l/m \to \infty$, so that the integral is finite when the cutoff is removed. Besides, we can write this square bracket as indicated in the second line of the following equation, which leads to a direct calculation of the leading term

$$\lim_{\Lambda\to\infty} (B - C) = \frac{g^2}{12\pi^2} \int_0^\infty \mathrm{d}l \left[\frac{8E_l^3}{m^4} - \frac{8l^3}{m^4} - \frac{12E_l}{m^2} + \frac{3}{E_l} + \frac{m^2}{E_l^3} \right]$$

$$= \frac{g^2}{12\pi^2 m^4} \int_0^\infty \mathrm{d}l \, \frac{\mathrm{d}}{\mathrm{d}l} \left[2lE_l^3 - 2l^4 - 3m^2 l E_l + \frac{m^4 l}{E_l} \right]$$

$$= \frac{g^2}{48\pi^2} . \tag{80}$$

The nonzero value of $B - C$ shows that the one-loop fermion self-energy introduces a Lorentz-violating term at small momenta even when the cut-off is made very large: Lorentz violation is suppressed not by a power of m/Λ, but only by a factor of the coupling. A convenient interpretation of this result uses the language of effective field theories. Observe that the Lorentz violation caused by the $B - C$ term is equivalent at the one-loop level to adding a term

$$-i\xi\bar{\psi}\gamma_\mu W^\mu W^\nu \partial_\nu \psi \tag{81}$$

to the Lagrangian. That is, we could obtain the same results in the $\Lambda \gg m$ region by dropping the $B - C$ term in the self-energy (72) and then adding this extra term to the Lagrangian. In this extra term $\xi = g^2/48\pi^2$, is the coefficient calculated above, and $W^\mu = (1, \mathbf{0})$ covariantly specifies the rest frame of the cutoff. Summarizing, the fermion self-energy contribution has the effect of feeding back terms of large unnatural size into the Lagrangian, thus producing a strong fine tuning problem when higher corrections are taken into account.

The Boson Self Energy [74]

Similar results can be obtained from the boson self energy. Since the internal lines are both fermionic in this case, a more symmetrical way of incorporating

the physical cutoff is required. This is accomplished by extending the fermionic propagators to

$$\frac{i}{\gamma k - m + i\epsilon} \longrightarrow \frac{i\, f(|\mathbf{k}|/\Lambda)}{\gamma k - m + \Xi(\mathbf{k}) + i\epsilon}\,. \tag{82}$$

The function $\Xi(\mathbf{k})$ takes into account the modified dispersion relations of the theory to zeroth order and its contribution will not be taken into account in the following.

Consider only the physical momentum cutoff function and calculate the boson self-energy $\Pi(E, \mathbf{p}, \Lambda)$, defined in terms of the full propagator $\Delta(E, \mathbf{p})$

$$\Delta(E, \mathbf{p}) = \frac{-i\, f(|\mathbf{k}|/\Lambda)}{p^2 - \mu^2 - \Pi(E, \mathbf{p}, \Lambda) - i\epsilon} \tag{83}$$

The one loop calculation is

$$\begin{aligned}\Pi(E, \mathbf{p}, \Lambda) = -\frac{ig^2}{16\pi^4} \int d^4k\ f(|\mathbf{k}|/\Lambda) f(|\mathbf{k}+\mathbf{p}|/\Lambda) \\ \times \frac{\mathrm{Tr}[(\gamma\cdot k + m)\,(\gamma\cdot(k+p)+m)]}{(k^2 - m^2 + i\epsilon)\,[(k+p)^2 - m^2 + i\epsilon]}\,.\end{aligned} \tag{84}$$

The result can be presented as an expansion in even powers of the momenta

$$\Pi(E, \mathbf{p}, \Lambda) = A + p^2 B + \eta_{RC} p^\mu p^\nu W_\mu W_\nu + \hat{\Pi}(E, \mathbf{p}, \Lambda)\,, \tag{85}$$

where $\hat{\Pi}$ is convergent when the regulator is removed so that it contributes only with Lorentz violating terms which are suppressed by powers of p^2/Λ^2. In the frame where the cutoff Λ is defined, $W_\mu = (1, \mathbf{0})$, we obtain

$$\eta_{RC} = \frac{1}{2}\left(\left(\frac{\partial^2 \Pi(p)}{\partial p^0 \partial p^0}\right)_{p=0} + \left(\frac{\partial^2 \Pi(p)}{\partial p^1 \partial p^1}\right)_{p=0}\right)\,, \tag{86}$$

from (85). The calculation of the required pieces starts from (84) and after some algebra produces

$$\left(\frac{\partial^2 \Pi(p)}{\partial p^0 \partial p^0}\right)_{p=0} = -\frac{g^2}{32\pi^2} \int_0^\infty dk\ k^4\ \frac{f^2}{E_k^5}\,, \tag{87}$$

$$\begin{aligned}\left(\frac{\partial^2\, \Pi(p)}{\partial p^i \partial p^j}\right)_{p=0} = -\delta^{ij}\frac{g^2}{48\pi^2} \left[\int_0^\infty k^2 dk\ \left(f f'' \frac{k^2}{\Lambda^2 E_k^3} + f f' \frac{k}{\Lambda E_k^5}\left[k^2 + 4m^2\right]\right.\right. \\ \left.\left. - f^2 \frac{1}{2E_k^7}\left[3k^4 + 10m^2k^2 - 3m^4\right]\right)\right]\,.\end{aligned} \tag{88}$$

Here $f = f(k/\Lambda)$ and the derivatives, denoted by primes, are with respect to the argument of f. As usual we have $E_k = \sqrt{k^2 + m^2}$. Each expression (87) and (88) is logarithmically divergent, but their combination (86) is finite. Thus we can estimate the leading contribution to η_{RC} by setting $f = 1$ inside the

integrals with no derivatives with respect to f. The terms containing derivatives are handled by assuming that the contributions of f', f'' are sharply localized in the region $x = k/\Lambda \approx 1$. This implies that we neglect the mass contributions $m^2 \ll k^2 \approx \Lambda^2$ in the corresponding integrals. Also, we set equal to zero the resulting boundary terms which include f' in the required integrations by parts. The final result is

$$\eta_{RC} = \frac{g^2}{12\pi^2}\left[1 + 2\int_0^\infty dx\, x\, f'(x)^2\right] , \tag{89}$$

which is estimated in the range $\eta_{RC} \geq 10^{-3}$ using standard model couplings. On the other hand, η_{RC} can be interpreted as a correction δc to the boson (photon) speed, for which extremely tight bounds exists

$$\frac{\eta_{RC}}{2} = \frac{\delta c}{c} = \leq 10^{-20}.$$

The two examples presented here show that radiative corrections in preferred frames associated to space granularity, modelled by the physical cutoff function (68), induce LIV contributions which are enormously increased with respect to the already established bounds. That is to say, they are not suppressed by factors $1/M$ as expected initially. In this way, a naturalness problem arises and a fine tuning problem appears when considering higher order radiative corrections. One possibility to guarantee the stability of dimension three and four LIV contributions is via the introduction of a custodial symmetry. This option has been explored using supersymmetry in [76].

An alternative way of dealing with radiative corrections in LIV processes can be found in [77].

4.3 Radiation in Lorentz Violating Electrodynamics

The observation of 100 MeV synchrotron radiation from the Crab Nebula has recently been used to impose stringent limits upon the parameters describing a modified electrodynamics embodied in Maxwell equations, which together with the corresponding coupled equations for the charges, get correction terms which are linear in the Planck length [15]. Such bounds were based on a set of very reasonable assumptions on how some of the standard results of synchrotron radiation extend to the Lorentz non-invariant situation. This caused some controversy in the literature [78]. Moreover, an assessment of such assumptions requires the introduction of specific dynamical models. One of them is the Myers–Pospelov (MP) effective theory at the classical level, which parameterizes LIV using dimension five operators [14]. A detailed description of synchrotron radiation in this model has been presented in [52], which we review in this subsection and that has motivated the general point of view summarized in Subsect. 4.4, given in more detail in [53].

The study of radiation in LIV electrodynamics constitutes an interesting problem on its own whose resolution will subsequently allow the use of additional

observational information to put bounds upon the correction parameters. For example we have in mind polarization measurements from cosmological sources. The case of gamma ray bursts has recently become increasingly relevant [79], although it is still at a controversial stage [80].

Our calculation of synchrotron radiation in modified electrodynamics rest heavily on previous work reported in [81–85].

A partial list of previous studies in electrodynamics incorporating LIV via dimension three and four operators is given in [86]

Synchrotron Radiation in the Myers–Pospelov Model

This model parameterizes LIV using dimension five operators both in the matter and electromagnetic sectors. There is also a preferred frame four velocity V^μ, which is not a dynamical field. As usual the model exhibits passive (observer) Lorentz covariance, which means that the fields and the four-velocity V^μ transform as tensors when going from one observer frame to another. On the other hand, in each frame we violate active (particle) Lorentz transformations; that is to say we have the non-dynamical physical field V^μ in the action, in analogy to the physics going on in the presence of an external magnetic field which violates active rotational invariance, for example.

The Charge Sector

The dynamics of a classical charged particle of mass μ can be obtained from the action for a scalar charged field. In this case the Myers–Pospelov action is

$$S_{MP} = \int d^4x \, \left[\partial_\mu \varphi^* \partial^\mu \varphi - \mu^2 \varphi^* \varphi + i\tilde{\eta}\, \varphi^* \left(V \cdot \partial \right)^3 \varphi \right] \, , \tag{90}$$

with the notation $V \cdot \partial = V^\mu \, \partial_\mu$. In momentum space, where we write $\varphi(x) = \varphi_0 \, \exp i(p^0 t - \mathbf{p} \cdot \mathbf{x})$ and in the reference frame where $V^\alpha = (1, \mathbf{0})$, the modified dispersion relation becomes

$$\left(p^0\right)^2 + \tilde{\eta} \left(p^0\right)^3 = \mathbf{p}^2 + \mu^2 \, . \tag{91}$$

To make contact with the results in [15] it is necessary to rewrite $\tilde{\eta} = -\eta/M$, $\eta < 0$, where η is a dimensionless constant and M is a mass scale characterizing the Lorentz symmetry breaking, which is usually, but not necessarily, identified with the Plank mass. The equation (91) is an exact relation in $\tilde{\eta}$. From here we obtain the Hamiltonian for a massive particle to second order in $\tilde{\eta}$

$$p^0 = H = \left(\mathbf{p}^2 + \mu^2\right)^{1/2} - \frac{1}{2}\tilde{\eta} \left(\mathbf{p}^2 + \mu^2\right) + \frac{5}{8}\tilde{\eta}^2 \left(\mathbf{p}^2 + \mu^2\right)^{3/2} + O(\tilde{\eta}^3) \, . \tag{92}$$

In the following we consider the interaction of a particle having charge q with a static constant magnetic field $\mathbf{B} = \boldsymbol{\nabla} \times \mathbf{A}$. The standard minimal coupling produces the Hamiltonian

$$H = \left[\left(\mathbf{p}-\frac{q}{c}\mathbf{A}\right)^2+\mu^2\right]^{1/2} - \frac{1}{2}\tilde{\eta}\left[\left(\mathbf{p}-\frac{q}{c}\mathbf{A}\right)^2+\mu^2\right] + \frac{5}{8}\tilde{\eta}^2\left[\left(\mathbf{p}-\frac{q}{c}\mathbf{A}\right)^2+\mu^2\right]^{3/2} , \tag{93}$$

to order $O(\tilde{\eta}^3)$. Here $c = 3\times 10^{10}\,\mathrm{cm\,s^{-1}}$ denotes the uncorrected velocity of light in vacuum. In the following we set $c = 1$. Observe that the dispersion relation (91) provides the exact inversion

$$(\mathbf{p} - q\mathbf{A})^2 = (1+\tilde{\eta}E)\,E^2 - \mu^2 , \tag{94}$$

with E being the energy of the particle. The Hamilton equations arising from (93) yield the acceleration

$$\ddot{\mathbf{r}} = \frac{q}{E}\left(1 - \frac{3}{2}\tilde{\eta}E + \frac{9}{4}\tilde{\eta}^2E^2\right)(\mathbf{v}\times\mathbf{B}) . \tag{95}$$

As in the usual case, this means that the magnitude $|\mathbf{v}|$ of the particle velocity is constant and that the projection of the trajectory in a plane perpendicular to $\mathbf{B}$ is a circular orbit with a Larmor frequency

$$\omega_0 = \frac{|q|B}{E}\left(1 - \frac{3}{2}\tilde{\eta}E + \frac{9}{4}\tilde{\eta}^2E^2\right) . \tag{96}$$

In general the motion is an helix with pitch angle (the angle between the velocity and the magnetic field) α. We restrict ourselves to the motion in the plane perpendicular to the magnetic field, i.e. $\alpha = \pi/2$. The solution to the equations of motion can be written as

$$\mathbf{r}(t) = \left(\frac{\beta}{\omega_0}\cos\omega_0 t,\ \frac{\beta}{\omega_0}\sin\omega_0 t,\ 0\right) , \tag{97}$$

where we emphasize that we are using the standard definition $\beta = |\mathbf{v}|/c$. The velocity is

$$\mathbf{v}(t) = (-\beta\sin\omega_0 t,\ \beta\cos\omega_0 t,\ 0) . \tag{98}$$

The equation (97) identifies $R = \beta/\omega_0$ as the Larmor radius of the orbit. The Lorentz factor γ is given by

$$\frac{1}{\gamma^2} = 1-\beta^2 = \frac{\mu^2}{E^2}\left[1 + 2\frac{\tilde{\eta}E^3}{\mu^2} - \frac{15}{4}\frac{\tilde{\eta}^2E^4}{\mu^2} + O(\tilde{\eta}^3)\right] , \tag{99}$$

where the range of energies to be considered is such that $\frac{\mu}{E} \ll 1$, $\tilde{\eta}E \ll 1$.

According to the preceding analysis, the current for a charged particle moving in the circular motion is

$$\mathbf{j}(t,\mathbf{r}) = q\delta^3(\mathbf{r}-\mathbf{r}(t))\,\mathbf{v}(t) , \tag{100}$$

where $\mathbf{r}(t)$ and $\mathbf{v}(t)$ are given in (97) and (98) respectively.

The Electromagnetic Sector

The corresponding action is [14]

$$S_{photon} = \int d^4x \left[-\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - 4\pi\, J^\mu A_\mu + \tilde{\xi}\, (V^\alpha F_{\alpha\delta})\, (V^\nu \partial_\nu)(V_\beta \tilde{F}^{\beta\delta}) \right] . \tag{101}$$

We choose to work in the rest frame $V^\mu = (1, \mathbf{0})$ where the modified Maxwell equations are ($c = 1$)

$$\begin{aligned} &\nabla \cdot \mathbf{B} = 0, \qquad && \nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0, \\ &\nabla \cdot \mathbf{E} = 4\pi\rho, \qquad && -\frac{\partial \mathbf{E}}{\partial t} + \nabla \times \mathbf{B} + \tilde{\xi} \partial_0 \left(-\nabla \times \mathbf{E} + \partial_0 \mathbf{B} \right)) = 4\pi \mathbf{j} \, . \end{aligned} \tag{102}$$

For future purposes it is convenient to define $\tilde{\xi} = \xi / M$, where ξ is a dimensionless parameter. Introducing the standard potential fields $A^\mu = (\phi, \mathbf{A})$ in the Coulomb gauge we have

$$\phi = -4\pi \frac{1}{\nabla^2} \rho, \tag{103}$$

$$\frac{\partial^2 \mathbf{A}}{\partial t^2} - \nabla^2 \mathbf{A} + 2\tilde{\xi} \nabla \times \frac{\partial^2 \mathbf{A}}{\partial t^2} = 4\pi \left(\mathbf{j} - \nabla \frac{1}{\nabla^2} \nabla \cdot \mathbf{j} \right) \equiv 4\pi\, \mathbf{j}_T \, , \tag{104}$$

where the electric and magnetic fields reduce to

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t}, \quad \mathbf{B} = \nabla \times \mathbf{A} \, , \tag{105}$$

in the radiation approximation.

The energy momentum tensor $T_{\mu\nu}$ for this modified electrodynamics is given by

$$T^0_0 = \frac{1}{8\pi} (\mathbf{E}^2 + \mathbf{B}^2) - \frac{\tilde{\xi}}{4\pi} \mathbf{E} \cdot \frac{\partial \mathbf{B}}{\partial t}, \tag{106}$$

$$\mathbf{S} = \frac{1}{4\pi} \mathbf{E} \times \mathbf{B} - \frac{\tilde{\xi}}{4\pi} \mathbf{E} \times \frac{\partial \mathbf{E}}{\partial t} \, , \tag{107}$$

which are exact expressions in $\tilde{\xi}$ and satisfy the usual conservation equation outside the sources.

To solve the equation of motion for $\mathbf{A}$, (104), we can go to the momentum space with the convention

$$F(t, \mathbf{r}) = \int \frac{d^4k}{(2\pi)^4} e^{-i\omega t + i \mathbf{k} \cdot \mathbf{r}} F(\omega, \mathbf{k}) \, . \tag{108}$$

The different types of Fourier transforms are specified by the corresponding arguments. For example, if $F(t, \mathbf{r})$ denotes the function in space-time, $F(\omega, \mathbf{r})$

denotes the Fourier transformed function to frequency space, while $F(\omega, \mathbf{k})$ denotes the Fourier transformed function to frequency and momentum spaces. In this way (104) reduces to

$$\left(-\omega^2 + k^2 - 2i\tilde{\xi}\omega^2\, \mathbf{k}\times\right) \mathbf{A}(\omega, \mathbf{k}) = 4\pi\, \mathbf{j}_T(\omega, \mathbf{k})\ . \tag{109}$$

This equation can be diagonalized using the circular polarization basis, with $\lambda = \pm 1$, giving

$$\left(-\omega^2 + k^2 \mp 2\tilde{\xi}\omega^2 k\right) \mathbf{A}^{\pm}(\omega, \mathbf{k}) = 4\pi \mathbf{j}_T^{\pm}(\omega, \mathbf{k})\ . \tag{110}$$

The components C_i^{λ} of any vector $\mathbf{C}$ in the polarization basis associated to the direction $\mathbf{k}$ are

$$C_i^{\lambda} = P_{ik}^{\lambda} C_k\,, \qquad P_{ik}^{\lambda} = \frac{1}{2}\left(\delta_{ik} - \hat{k}_i\hat{k}_k + \lambda i\epsilon_{ijk}\hat{k}_j\right)\ . \tag{111}$$

The simplest way to proceed is by introducing the total retarded Green function

$$[G_{ret}(\omega, \mathbf{k})]_{ik} = \sum_{\lambda} P_{ik}^{\lambda}\ \left.\frac{1}{k^2 - \lambda 2\tilde{\xi}\omega^2 k - \omega^2}\right|_{\omega\to\omega+i\epsilon}\ . \tag{112}$$

and to calculate

$$[G_{ret}(\omega, \mathbf{r}-\mathbf{r}')]_{ik} = \int \frac{d^3\mathbf{k}}{(2\pi)^3} e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} [G_{ret}(\omega, \mathbf{k})]_{ik}\ . \tag{113}$$

After some rearrangements the final integration over dk can be performed by the method of residues in the complex plane. Finally one can identify back the polarization components of the total Green function as

$$G_{ret}^{\lambda}(\omega, \mathbf{r}-\mathbf{r}') = \frac{1}{4\pi R}\frac{n_\lambda(z)}{\sqrt{1+z^2}}\, e^{in_\lambda(z)\omega R}\ , \tag{114}$$

where $R = |\mathbf{r}-\mathbf{r}'|$. Here we have introduced the polarization-dependent refraction index $n_\lambda(z)$

$$n_\lambda(z) = \sqrt{1+z^2} + \lambda z, \quad z = \tilde{\xi}\omega\ . \tag{115}$$

In this way, the fields $\mathbf{A}^{\lambda}$ in (110) have well defined phase velocities $v_\lambda = 1/n_\lambda(z)$ and this situation can be described as the propagation of photons in a dispersive birefringent medium.

The Green functions (114) determine the corresponding potentials with the standard replacements $1/R \simeq 1/|\mathbf{r}| \equiv 1/r$ in the denominator, together with the following expansion of the phase $n(\lambda z)\omega R$

$$n_\lambda(z)\omega\,|\mathbf{r}-\mathbf{r}'| \simeq \omega r\left[1 - \frac{\hat{\mathbf{n}}\cdot\mathbf{r}'}{r} + \lambda\tilde{\xi}\omega - \lambda\tilde{\xi}\omega\frac{\hat{\mathbf{n}}\cdot\mathbf{r}'}{r} + \frac{1}{2}\left(\frac{r'}{r}\right)^2\right]\ , \tag{116}$$

where $\hat{\mathbf{n}} = \mathbf{r}/r$ is the direction of observation. Notice that we are interested in the radiation approximation of the phase (116), which means that the subdominant terms of order $(r'/r)^2$ or higher are neglected. Consistency demands that the terms proportional to the LIV parameter $\tilde{\xi}$ are larger than the neglected one in order to properly include them in this phase. Our general results are presented in this full far-field approximation.

Using (114) we finally get

$$\mathbf{A}^{\lambda}(\omega, \mathbf{r}) = \frac{1}{r} \frac{n_{\lambda}(z)}{\sqrt{1+z^2}} e^{in_{\lambda}(z)\omega r} \mathbf{j}^{\lambda}(\omega, \mathbf{k}_{\lambda}) \tag{117}$$

in the radiation approximation. The fields $\mathbf{A}^{+}(\omega, \mathbf{r})$ and $\mathbf{A}^{-}(\omega, \mathbf{r})$ correspond to right and left circular polarization respectively. Let us emphasize that the momenta

$$\mathbf{k}_{\lambda} = n_{\lambda}(z)\ \omega\ \hat{\mathbf{n}} \tag{118}$$

in (117) are fixed in terms of the frequency and the direction of observation. The full vector potential is given by the superposition $\mathbf{A}(\omega, \mathbf{r}) = \mathbf{A}^{+}(\omega, \mathbf{r}) + \mathbf{A}^{-}(\omega, \mathbf{r})$. Hence the electric and magnetic fields are

$$\begin{aligned} \mathbf{B}(\omega, \mathbf{r}) &= \frac{1}{r} \frac{\omega}{\sqrt{1+z^2}} \sum_{\lambda=\pm} \lambda\, n_{\lambda}^2(z)\, e^{in_{\lambda}(z)\omega r}\, \mathbf{j}^{\lambda}(\omega, \mathbf{k}_{\lambda}), \\ \mathbf{E}(\omega, \mathbf{r}) &= \frac{1}{r} \frac{i\omega}{\sqrt{1+z^2}} \sum_{\lambda=\pm} n_{\lambda}(z)\, e^{in_{\lambda}(z)\omega r}\, \mathbf{j}^{\lambda}(\omega, \mathbf{k}_{\lambda})\ . \end{aligned} \tag{119}$$

Note that, contrary to the standard case where $\hat{\mathbf{n}} \times \mathbf{E} \propto \mathbf{B}$, here we have

$$\hat{\mathbf{n}} \times \mathbf{E}(\omega, \mathbf{r}) = \frac{1}{\sqrt{1+z^2}} \left[\mathbf{B}(\omega, \mathbf{r}) + iz\mathbf{E}(\omega, \mathbf{r})\right]\ . \tag{120}$$

The angular distribution of the power spectrum is defined as

$$\frac{d^2P(T)}{d\omega d\Omega}\ , \tag{121}$$

where $P(T)$ is the radiated power at time T into the solid angle $d\Omega$. We can compute the total energy emitted in terms of the Poynting vector (107)

$$E = \int_{-\infty}^{\infty} dt\ \mathbf{n} \cdot \mathbf{S}(t, \mathbf{r})\ r^2 d\Omega\ . \tag{122}$$

This last expression can be rewritten introducing the Fourier transform of the Poynting vector,

$$E = \int_0^{\infty} d\omega \int d\Omega\ \frac{d^2E}{d\Omega d\omega} = \int_0^{\infty} \frac{d\omega}{2\pi} \left[\mathbf{n} \cdot \mathbf{S}(\omega, \mathbf{r}) + \mathbf{n} \cdot \mathbf{S}(-\omega, \mathbf{r})\right] r^2 d\Omega\ , \tag{123}$$

and allows us to obtain the angular distribution of the energy spectrum

$$\frac{d^2E}{d\Omega d\omega} = \frac{r^2}{2\pi}\left[\mathbf{n}\cdot\mathbf{S}(\omega,\mathbf{r}) + \mathbf{n}\cdot\mathbf{S}(-\omega,\mathbf{r})\right] \ , \tag{124}$$

from where the angular distribution of the power spectrum can be identified as

$$\frac{d^2E}{d\Omega d\omega} = \int_{-\infty}^{+\infty} dT \ \frac{d^2P(T)}{dwd\Omega} \ . \tag{125}$$

Writing the Poynting vector (107) in terms of the polarized potentials we obtain the intermediate result

$$\frac{d^2E}{d\Omega d\omega} = \frac{r^2\omega^2}{4\pi^2}\sqrt{1+z^2}\left[\mathbf{A}_-(-\omega,\mathbf{r})\cdot\mathbf{A}_+(\omega,\mathbf{r}) + \mathbf{A}_-(\omega,\mathbf{r})\cdot\mathbf{A}_+(-\omega,\mathbf{r})\right] \ . \tag{126}$$

Next we express the products $\mathbf{A}_\mp(-\omega,\mathbf{r})\cdot\mathbf{A}_\pm(\omega,\mathbf{r})$ in terms of the current $\mathbf{j}(\omega,\ \mathbf{k})$ via the relation (117). Using the properties of the fields in the polarization basis, together with the general relation $j_k(-\omega,-\mathbf{k}) = j_k^*(\omega,\mathbf{k})$, we obtain

$$\begin{aligned}\frac{d^2E}{d\Omega d\omega} = \frac{1}{4\pi^2}\frac{\omega^2}{1+z^2}\Big[& n_+^2(z)\ j_k^*(\omega,\mathbf{k}_+)\ P_{kr}^+\ j_r(\omega,\mathbf{k}_+) \\ & + n_-^2(z)\ j_k^*(\omega,\mathbf{k}_-)\ P_{kr}^-\ j_r(\omega,\mathbf{k}_-)\ \Big] \ .\end{aligned} \tag{127}$$

In order to identify the angular distribution of the power spectrum we need to introduce the time dependence via the corresponding inverse Fourier transform. Each contribution in (127) is of the type

$$C(\omega) = j_k^*(\omega,\mathbf{k})\ X_{kr}\ j_r(\omega,\mathbf{k}) = \int_{-\infty}^{+\infty} dt\ dt' e^{-i\omega(t-t')} j_k^*(t,\mathbf{k})\ X_{kr}\ j_r(t',\mathbf{k}) \ . \tag{128}$$

Changing to new time variables $\tau = t - t'$ and $T = (t+t')/2$ we get

$$C(\omega) = \int_{-\infty}^{+\infty} dT \int_{-\infty}^{\infty} d\tau e^{-i\omega\tau} j_k^*(T+\tau/2,\mathbf{k})\ X_{kr}\ j_r(T-\tau/2,\mathbf{k}) \ . \tag{129}$$

Inserting this last relation in (127) and comparing with (125) we obtain the final expression for the angular distribution of the radiated power spectrum

$$\begin{aligned}\frac{d^2P(T)}{d\omega d\Omega} = & \frac{1}{4\pi^2}\frac{\omega^2}{\sqrt{1+z^2}}\int_{-\infty}^{\infty} d\tau\ e^{-i\omega\tau} \\ & \times \sum_\lambda n_\lambda^2(z) j_k^*(T+\tau/2,\mathbf{k}_\lambda)\, P_{kr}^\lambda\ j_r(T-\tau/2,\mathbf{k}_\lambda) \ ,\end{aligned} \tag{130}$$

as the sum of the contributions of both circular polarizations.

Synchrotron Radiation

This corresponds to the choice (98) for the velocity of the current (100) in the general expression (130). Following the method of [82] the time-averaged angular distribution of the radiated power spectrum is

$$\left\langle \frac{d^2P(T)}{d\omega d\Omega} \right\rangle = \sum_{\lambda=+,-} \sum_{m=0}^{\infty} \delta(\omega - \omega_m) \frac{dP_{m,\lambda}}{d\Omega}, \quad \omega_m = m\omega_0, \quad z_m = \tilde{\xi}\omega_m,$$
$$\frac{dP_{m,\lambda}}{d\Omega} = \frac{\omega_m^2 q^2}{4\pi} \frac{1}{\sqrt{1+z_m^2}} \left[\lambda\beta n_\lambda(z_m) J'_m(W_{\lambda m}) + \cot\theta\, J_m(W_{\lambda m})\right]^2 , \tag{131}$$

written as a sum over the contribution of the harmonics ω_m. Here, the average $\langle\rangle$ is taken with respect to the macroscopic time T and J_m, J'_m denote the Bessel functions and their derivatives respectively. The argument of the Bessel functions is $W_{\lambda m} = m\, n_\lambda(z_m)\beta \sin\theta$.

We also have calculated the total averaged and integrated power radiated into the m-th harmonic

$$P_m = \frac{q^2\beta^2\omega_m}{R\sqrt{1+z_m^2}} \sum_{\lambda=\pm} n_\lambda(z_m) \left[J'_{2m}(2m\, n_\lambda(z_m)\beta) - \frac{1-\beta^2 n_\lambda^2(z_m)}{2\beta^2 n_\lambda^2(z_m)} \int_0^{2mn_\lambda(z_m)\beta} dx\, J_{2m}(x)\right] , \tag{132}$$

which clearly indicates the contribution of each polarization $P_{\lambda m}$. The above result is exact in z_m and the parity-violating contribution has vanished after the angular integration.

In Table 3 we present a rough estimation of the relevant parameters associated with some observed cosmological objects. There r [l.y.] is the distance of the object to the earth, γ is the Lorentz factor of the charged particles at the cutoff frequency, B [Gauss] is the average magnetic field producing the synchrotron radiation, ω_c [GeV] is the cut-off frequency and ω_0 [GeV] is the Larmor frequency. In all cases the cut-off frequency ω_c has been estimated from the radiation spectrum fitted by a synchrotron model in the corresponding reference. This, together with the magnetic field B allows us to estimate the Lorentz factor

$$\gamma = 2.36 \times 10^8 \sqrt{\frac{\omega_c[\mathrm{GeV}]}{B[\mathrm{Gauss}]} \frac{M}{m_e}} , \tag{133}$$

where M is the mass of the charged particle. From the above we further obtain the Larmor frequency

Table 3. Data of some relevant astrophysical objects

Object	$r\,(l.y.)$	γ	$B\ (Gauss)$	$\omega_c(GeV)$	$\omega_0(GeV)$	m	m/γ
CRAB	10^4	10^9	10^{-3}	10^{-1}	10^{-30}	10^{29}	10^{20}
$(Mkn\,501)_p$	10^8	10^{11}	10^2	10^4	10^{-29}	10^{33}	10^{22}
$(Mkn\,501)_e$	10^8	10^{11}	10^{-1}	10^4	10^{-29}	10^{33}	10^{22}
$GRB\,021006$	10^{10}	10^5	10^4	10^{-3}	10^{-18}	10^{15}	10^{10}

$$\omega_0[GeV] = 0.6 \times 10^{-17} \left(\frac{2m_e}{\gamma M}\right) B[Gauss] \; . \tag{134}$$

In the case of CRAB Nebula we adopt the typical values given in [87]. For Mkn 501 we consider two possible models for synchrotron radiation where the emitter particles are either protons [88] or electrons [89]. In the latter case we use the radius of the orbit $r' = 1.5 \times 10^{15}\,\mathrm{cm} = 1/\omega_0$ and the magnetic field to obtain the Lorentz factor. Finally we consider the GRB021206 ($z \approx 1.25$). According to [90] this object has a very compact core with a radius of the order of 1 km and a magnetic field $\approx 10^{12}$ Gauss. The synchrotron emission region is about 10^8 km from the core [91], so that we estimate the magnetic field to be 10^4 Gauss using the transport law $B\,r = const.$. From [92] we take the cut-off frequency to be $\omega_c = 1$ MeV.

As indicated in Table 3., the radiation of interest is dominated by very high harmonics $10^{15} \leq m \leq 10^{30}$ exhibiting also large ratios of m/γ, typically in the range $10^{10} \leq m/\gamma \leq 10^{22}$. The corresponding values for γ imply also $\beta \approx 1$. In this way, the high harmonics present in the synchrotron radiation of these astrophysical sources together with the values of the γ factors of the radiating charges highlight the relevance of the large m and large γ limit, with the constraint $(m/\gamma)^2 \gg 1$, to study the induced Lorentz violating effects.

In a similar way to the standard case [83], we obtain

$$P_{\lambda m} = \frac{q^2 m\omega_0}{\sqrt{3}\pi R} \frac{1}{1+n^2(\lambda z_m)} \left\{ \int_{m/\tilde{m}_{\lambda c}}^{\infty} dx \left(\frac{3}{2\tilde{m}_{\lambda c}}\right)^{2/3} K_{5/3}(x) \right.$$
$$\left. -2\left(\frac{3}{2\tilde{m}_{\lambda c}}\right)^{4/3} K_{2/3}\left(\frac{m}{\tilde{m}_c}\right) \right\} \; ,$$

for the integrated power in the m^{th} harmonic. Here $K_{p/q}$ denote the Macdonald functions (Bessel functions of fractional order). A first consequence of this approximation is the appearance of the cutoff frequency $\omega_{\lambda c} = \tilde{m}_{\lambda c}\,\omega_0$ with

$$\tilde{m}_{\lambda c} = \frac{3}{2}\left[1 - \beta^2(E)\, n_\lambda^2(z_m)\right]^{-3/2} \; . \tag{135}$$

This name arises because for $m > \tilde{m}_{\lambda c}$ the total power decreases as

$$P_{\lambda m} \approx \exp(-m/\tilde{m}_{\lambda c}) \; . \tag{136}$$

Within the same large-m approximation, the integrated power in the m-th harmonic can be expanded to second order in $\tilde{\xi}$ yielding

$$P_m = \frac{q^2\omega}{\sqrt{3}\pi R\gamma^2} \left[\frac{m_c}{m}\kappa\left(\frac{m}{m_c}\right) - \frac{2}{\gamma^2} K_{2/3}\left(\frac{m}{m_c}\right) \right.$$
$$\left. +2\left(\frac{\tilde{\xi}\, m\omega\beta}{\gamma}\right)^2 K_{2/3}\left(\frac{m}{m_c}\right) \right] \; , \tag{137}$$

where $m_c = 3\gamma^3/2$ and $\kappa(x) = x\int_x^\infty dy\, K_{5/3}(y)$ is the so called bremsstrahlung function.

Let us notice the appearance of the combination $\tilde{\xi}\,\omega m/\gamma = \xi(\omega/M_P)(m/\gamma)$ as the expansion parameter in (137). Here we take $M = M_P$. As can be seen from Table 3 this is not necessarily a small number, which signals the possibility that such corrections might be relevant in setting bounds upon $\tilde{\xi}$. This rather unexpected effect is due to the amplifying factor m/γ. Similar results have been obtained in calculations of the synchrotron radiation spectra in the context of non-commutative electrodynamics [93].

Another possibility for observable effects due to $\tilde{\xi}$ is to look at the averaged degree of circular polarization

$$\Pi_\odot = \frac{\langle P_+(\omega) - P_-(\omega)\rangle}{\langle P_+(\omega) + P_-(\omega)\rangle}\,, \tag{138}$$

where $P_\lambda(\omega)$ is the total power distribution per unit frequency and polarization λ, so that $P_\lambda(\omega) = P_{m\lambda}/\omega_0$. The average here is calculated with respect to an energy distribution of the relativistic electrons, which we take to be $N(E)dE = CE^{-p}dE$, in some energy range $E_1 < E < E_2$, chosen as $E_1 = 0$ and $E_2 \to \infty$ for simplicity. The result is

$$\Pi_\odot = \tilde{\xi}\omega\left(\frac{\mu\omega}{qB}\right)\Pi(p)\,, \tag{139}$$

$$\Pi(p) = \frac{(p-3)\,(3p-1)}{3\,(3p-7)}\,\frac{(p+1)}{(p-1)}\,\frac{\Gamma\left(\frac{p}{4}+\frac{13}{12}\right)\Gamma\left(\frac{p}{4}+\frac{5}{12}\right)}{\Gamma\left(\frac{p}{4}+\frac{19}{12}\right)\Gamma\left(\frac{p}{4}+\frac{11}{12}\right)}, \quad p > 7/3\,. \tag{140}$$

Again, we have the presence of an amplifying factor in (139), given by $(\mu\omega/qB)$, which is independent of the form of $\Pi(p)$ and not necessarily a small number. An estimation of this factor in the zeroth-order approximation ($\tilde{\xi} = 0 = \tilde{\eta}$), which is appropriate in (139), yields $(\mu\omega/qB) = \omega/(\omega_0\gamma) = m/\gamma$. The expression (139) is analogous to the well-known average of the degree of linear polarization $\Pi_{LIN} = (p+1)/(p+7/3)$, under the same energy distribution for the electrons [94].

Finally in Table 4 we have estimated the contributions to the different pieces in the phase (116) of the Green function, for each astrophysical object. Here we have set $\xi = 1$ and $M = M_P$. The extreme case is

Table 4. The far field approximation

Object	r'/r	ω_c/M_P	$(r'/r)(\omega_c/M_P)$	$(r'/r)^2$
CRAB	10^{-6}	10^{-20}	10^{-26}	10^{-12}
$Mkn\,501$	10^{-11}	10^{-15}	10^{-26}	10^{-22}
$GRB\,021006$	10^{-24}	10^{-22}	10^{-46}	10^{-48}

$$\left|\tilde{\xi}\omega\right| \frac{r'}{r} < \left|\tilde{\xi}\omega\right| < \left(\frac{r'}{r}\right)^2 , \tag{141}$$

where all the dependence on $\tilde{\xi}$ is negligible in the phase, which reduces to

$$(\lambda z)\omega \left|\mathbf{r} - \mathbf{r}'\right| \simeq \omega(r - \hat{n} \cdot \mathbf{r}') . \tag{142}$$

This corresponds to the CRAB nebulae case, where the assumptions made in [15] are readily verified. A detailed discussion of synchrotron radiation in the Myers–Pospelov model can be found in the third reference [52].

4.4 General Point of View of LIV Radiation

Three paradigmatic examples of Lorentz violating electrodynamics are given by the effective theories proposed by Gambini and Pullin (GP) [30], Ellis et al. (EMN) [28], and Myers and Pospelov (MP) [14]. They can be written in the general form of Maxwell equations

$$\nabla \cdot \mathbf{D} = 4\pi\rho, \qquad \nabla \cdot \mathbf{B} = 0, \tag{143}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad \nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} + 4\pi\mathbf{j} , \tag{144}$$

with corresponding constitutive relations

$$\mathbf{D} = \mathbf{D}(\mathbf{E}, \mathbf{B}), \quad \mathbf{H} = \mathbf{H}(\mathbf{E}, \mathbf{B}) , \tag{145}$$

which we next present in detail for each case, after reviewing the corresponding equations. Let us recall that the above equations (143) and (144) imply charge conservation $\partial\rho/\partial t + \nabla \cdot \mathbf{j} = 0$, independently of the constitutive equations (145). In an abuse of notation we have denoted by $\tilde{\xi}$ the electromagnetic LIV parameter for all models in the sequel.

Gambini-Pullin Electrodynamics

The Maxwell equations for this case are

$$\nabla \cdot \mathcal{B} = 0, \qquad \nabla \times \left(\mathcal{E} + 2\tilde{\xi}\nabla \times \mathcal{E}\right) + \frac{\partial \mathcal{B}}{\partial t} = 0 , \tag{146}$$

$$\nabla \cdot \mathcal{E} = 4\pi\rho, \qquad \nabla \times \left(\mathcal{B} + 2\tilde{\xi}\nabla \times \mathcal{B}\right) - \frac{\partial \mathcal{E}}{\partial t} = 4\pi\mathbf{j} , \tag{147}$$

where the electric and magnetic fields are identified from the homogeneous equation as

$$\mathbf{E} = \mathcal{E} + 2\tilde{\xi}\nabla \times \mathcal{E}, \qquad \mathbf{B} = \mathcal{B} . \tag{148}$$

From the inhomogeneous equations we obtain

$$\mathbf{D} = \mathcal{E}, \qquad \mathbf{H} = \mathcal{B} + 2\tilde{\xi}\nabla \times \mathcal{B} , \tag{149}$$

which together with the constitutive relations

$$\mathbf{D} + 2\tilde{\xi}\nabla \times \mathbf{D} = \mathbf{E}, \qquad \mathbf{H} = \mathbf{B} + 2\tilde{\xi}\nabla \times \mathbf{B}\ , \tag{150}$$

leave the equations in the required form. These equations define the corresponding functions stated in (145). In momentum space we have

$$\mathbf{D} = \frac{1}{1 + 4\tilde{\xi}^2 k^2}\left(\mathbf{E} - 2i\tilde{\xi}\mathbf{k} \times \mathbf{E} + 4\tilde{\xi}^2\,(\mathbf{k}\cdot\mathbf{E})\,\mathbf{k}\right), \qquad \mathbf{H} = \mathbf{B} + 2i\tilde{\xi}\mathbf{k} \times \mathbf{B}\ . \tag{151}$$

The admixture of vectors and axial vectors in the constitutive relations precludes the parity violation exhibited by the model, together with the presence of birefringence.

Ellis et al. Electrodynamics

In this case the modified Maxwell equations are

$$\nabla \cdot \mathbf{B} = 0, \qquad \nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0\ , \tag{152}$$

$$\nabla \cdot \mathbf{E} + \mathbf{u} \cdot \frac{\partial \mathbf{E}}{\partial t} = 4\pi\rho_{\,\mathrm{eff}} = 4\pi(\rho - \mathbf{u}\cdot\mathbf{j})\ , \tag{153}$$

$$\begin{aligned}\nabla \times \mathbf{B} - \left(1 - \mathbf{u}^2\right)\frac{\partial \mathbf{E}}{\partial t} + \mathbf{u} \times \frac{\partial \mathbf{B}}{\partial t} + (\mathbf{u}\cdot\nabla)\,\mathbf{E} &= 4\pi\mathbf{j}_{\,\mathrm{eff}} \\ &= 4\pi(\mathbf{j} + \mathbf{u}(\rho - \mathbf{u}\cdot\mathbf{j}))\ ,\end{aligned} \tag{154}$$

which can be written in the form (143-144) via the constitutive relations [28]

$$\begin{aligned}&\mathbf{H} = \mathbf{B} - f(\omega)\mathbf{k} \times \mathbf{E}\ , \\ &\mathbf{D} = \left(1 - f^2(\omega)k^2\right)\mathbf{E} + f^2(\omega)\mathbf{k}\,(\mathbf{k}\cdot\mathbf{E}) - f(\omega)\mathbf{k} \times \mathbf{B}\ ,\end{aligned} \tag{155}$$

where we have assumed that $\mathbf{u} = f(\omega)\mathbf{k}$ in momentum space. Taking $\mathbf{u}$ as a vector, this model conserves parity and shows no birefringence.

Myers–Pospelov Electrodynamics

This case corresponds to the equations (102). From the last one we can infer the constitutive relations

$$\mathbf{H} = \mathbf{B} - \tilde{\xi}\partial_0\mathbf{E}, \qquad \mathbf{D} = \mathbf{E} - \tilde{\xi}\partial_0\mathbf{B}\ , \tag{156}$$

which produce

$$\nabla \cdot \mathbf{E} = \nabla \cdot \mathbf{D}\ , \tag{157}$$

leaving the third (102) in desired form. Similarly to the GP case, this model violates parity. In momentum space (156) become

$$\mathbf{H} = \mathbf{B} + i\tilde{\xi}\omega\mathbf{E}, \qquad \mathbf{D} = \mathbf{E} + i\tilde{\xi}\omega\mathbf{B}\ . \tag{158}$$

The above constitutive relations in the three representative models involve linear relations among the fields and can be summarized, in momentum space, as the local relations

$$D_i(\omega,\mathbf{k}) = \alpha_{ij}(\omega,\mathbf{k})E_i(\omega,\mathbf{k}) + \rho_{ij}(\omega,\mathbf{k})B_j(\omega,\mathbf{k}) \ ,$$
$$H_i(\omega,\mathbf{k}) = \beta_{ij}(\omega,\mathbf{k})B_i(\omega,\mathbf{k}) + \sigma_{ij}(\omega,\mathbf{k})E_j(\omega,\mathbf{k}) \ , \tag{159}$$

where the corresponding momentum dependent coefficients can be read from the equations (151), (155), and (158). Equations (159) are the most general expressions in which any pair of linear constitutive relations can be ultimately written, which allow to express the fields $\mathbf{D},\mathbf{H}$ in terms of $\mathbf{E},\mathbf{B}$.

Parameterization of the Constitutive Relations

Let us consider Maxwell equations in momentum space

$$\mathbf{k}\cdot\mathbf{B}\,(\omega,\mathbf{k}) = 0, \qquad \mathbf{k}\times\mathbf{E}\,(\omega,\mathbf{k}) = \omega\mathbf{B}\,(\omega,\mathbf{k}) \ , \tag{160}$$
$$i\mathbf{k}\cdot\mathbf{D}\,(\omega,\mathbf{k}) = 4\pi\rho\,(\omega,\mathbf{k})\,, \quad i\mathbf{k}\times\mathbf{H}\,(\omega,\mathbf{k}) = -i\omega\mathbf{D}\,(\omega,\mathbf{k}) \ + 4\pi\mathbf{j}\,(\omega,\mathbf{k}) \ . \tag{161}$$

Here we discuss the vacuum situation where the non trivial constitutive relations arise because of LIV effects. Let us take into account corrections up to second order in the LIV parameter $\tilde{\xi}$ and let us assume that we are in a Lorentz frame where we demand invariance under rotations. This would correspond to the rest frame $V^\mu = (1,\mathbf{0})$ in the MP model, for example. We can always go to an arbitrary frame by means of a passive (observer) Lorentz transformation. In this way we have the general expressions

$$\alpha_{ij} = \alpha_0\delta_{ij} + i\alpha_1\tilde{\xi}\epsilon_{irj}k_r + \alpha_2\tilde{\xi}^2k_ik_j, \quad \rho_{ij} = \rho_0\delta_{ij} + i\rho_1\tilde{\xi}\epsilon_{irj}k_r + \rho_2\tilde{\xi}^2k_ik_j \ ,$$
$$\beta_{ij} = \beta_0\delta_{ij} + i\beta_1\tilde{\xi}\epsilon_{irj}k_r + \beta_2\tilde{\xi}^2k_ik_j, \quad \sigma_{ij} = \sigma_0\delta_{ij} + i\sigma_1\tilde{\xi}\epsilon_{irj}k_r + \sigma_2\tilde{\xi}^2k_ik_j \ , \tag{162}$$

where $\alpha_A,\beta_A,\rho_A,\sigma_A, A = 0,1,2$, are scalar functions depending only upon ω, $k = |\mathbf{k}|$, and $\tilde{\xi}$. The property $\mathbf{k}\cdot\mathbf{B} = 0$ sets $\beta_2 = \rho_2 = 0$ effectively. In vector notation we then have

$$\mathbf{D} = \left(\alpha_0 + \alpha_2k^2\tilde{\xi}^2\right)\mathbf{E} + \left(\rho_0 + i\alpha_1\omega\tilde{\xi}\right)\mathbf{B} + i\left(\rho_1 - i\alpha_2\omega\tilde{\xi}\right)\tilde{\xi}\ \mathbf{k}\times\mathbf{B} \ ,$$
$$\mathbf{H} = \left(\sigma_0 + \sigma_2k^2\tilde{\xi}^2\right)\mathbf{E} + \left(\beta_0 + i\sigma_1\omega\tilde{\xi}\right)\mathbf{B} + i\left(\beta_1 - i\sigma_2\omega\tilde{\xi}\right)\tilde{\xi}\ \mathbf{k}\times\mathbf{B} \ , \tag{163}$$

where we have used the second (160) together with $(\mathbf{k}\cdot\mathbf{E})\,\mathbf{k} = \omega\,(\mathbf{k}\times\mathbf{B}) + k^2\mathbf{E}$.

The Generalized Maxwell Equations [53]

Next we substitute (163) in (161) to obtain the corresponding equations for $\mathbf{E}$ and $\mathbf{B}$. The result is

$$i\left(\alpha_0+\alpha_2 k^2\tilde{\xi}^2\right)(\mathbf{k}\cdot\mathbf{E})=4\pi\rho, \tag{164}$$

$$i\left(\alpha_0+\alpha_2 k^2\tilde{\xi}^2\right)\omega\mathbf{E}+i\left[\beta_0+i\left(\sigma_1+\rho_1\right)\omega\tilde{\xi}+\alpha_2\tilde{\xi}^2\omega^2\right]\mathbf{k}\times\mathbf{B} +i\left[\left(\sigma_0+\rho_0\right)\omega+i\left(\alpha_1\omega^2-\beta_1 k^2\right)\tilde{\xi}\right]\mathbf{B}=4\pi\mathbf{j}\left(\omega,\mathbf{k}\right)\ . \tag{165}$$

It is convenient to rewrite the inhomogeneous equations in the compact form

$$iP\left(\mathbf{k}\cdot\mathbf{E}\right)=4\pi\rho\ , \tag{166}$$

$$i\omega P\,\mathbf{E}+iQ\,\mathbf{k}\times\mathbf{B}+R\,\mathbf{B}\ =4\pi\mathbf{j}\left(\omega,\mathbf{k}\right)\ , \tag{167}$$

by defining

$$P=\alpha_0+\alpha_2\tilde{\xi}^2k^2,\qquad Q=\beta_0+i\left(\sigma_1+\rho_1\right)\omega\tilde{\xi}+\alpha_2\tilde{\xi}^2\omega^2,$$
$$R=\left(\beta_1k^2-\alpha_1\omega^2\right)\tilde{\xi}+i\left(\sigma_0+\rho_0\right)\omega\ . \tag{168}$$

Now we have only three independent functions which depend on ω and k. The inhomogeneous equations (166) and (167) can be solved by introducing the standard electromagnetic potentials in the Coulomb gauge. The radiation approximation is described in terms of the vector potential $\mathbf{A}^\lambda$ only, which will exhibit polarizations $\lambda=\pm1$ in the general case and that satisfies the equation

$$\left[Qk^2-P\omega^2+\lambda k\,R\right]\mathbf{A}^\lambda=4\pi\mathbf{j}_T^\lambda\ . \tag{169}$$

Here $\mathbf{j}_T^\lambda$ is the transverse part of the polarized current. The Green function $G^\lambda(\omega,k)$ corresponding to the operator in the LHS of (169) satisfies in general a cubic equation in k, which determines the dispersion relations for the propagating photon. This means that we have three poles $k_\lambda(\omega)$. In fact one of the poles is due to the β_1 factor in the expression for R in (168) and it is located at $k\sim\xi^{-1}$. Therefore its contribution can be neglected in our effective theory valid for $k\ll\xi^{-1}$. The remaining two poles will deviate little from their position when $\tilde{\xi}=0$ and will allow us to characterize the propagation mode corresponding to each polarization λ by a refraction index $n_\lambda(\omega)$ to be read from the appropriate dispersion relation in such a way that $n_\lambda(\omega)=k_\lambda(\omega)/\omega$. The general form of the polarized Green function will be

$$G^\lambda(\omega,\mathbf{r})=\frac{1}{4\pi r}F^\lambda(\omega)e^{i\omega n_\lambda(\omega)r}\ . \tag{170}$$

Notice that from the birefringent case we can go to the non-birefringent one by taking $n_+(\omega)=n_-(\omega)=n(\omega)$, in which case $F^+(\omega)=F^-(\omega)=F(\omega)$.

In the following we make explicit some general properties that must be satisfied by any electrodynamics characterized by a Green function of the type (170). The reality of the electric and magnetic fields leads to the condition

$$\left[G^+\left(\omega,\mathbf{r}\right)\right]^*=G^-\left(-\omega,\mathbf{r}\right)\ . \tag{171}$$

This implies the relations

$$n_+^*(\omega) = n_-(-\omega), \qquad [F^+(\omega)]^* = F^-(-\omega)\ . \tag{172}$$

For a birefringent medium the real and imaginary parts of the refraction index for circular polarization components can contain both ω-even and ω-odd terms, provided that they satisfy the first (172). In the case of a non-birefringent medium the real part of the refraction index must be even in ω, while the imaginary part must be odd. We can see that the refraction indices for the Myers–Pospelov theory, (115), satisfy these requirements. A detailed discussion of this generalized point of view is under preparation [95].

Acknowledgements

The author acknowledges C. Lämmerzahl for his invitation to participate in the 339th *WE-Heraeus-Seminar: Special Relativity, will it survive the next 100 years?*, for his invitation as well as for his wonderful hospitality at Potsdam. Partial support from the projects CONACYT-México-40745-F and DGAPA-UNAM-IN104503-3; as well as the hospitality of the George P. and Cynthia W. Mitchell Institute for Fundamental Physics, Texas A&M University are also gratefully acknowledged.

References

1. For reviews see for example: C. Rovelli: Loop quantum gravity, *Livings Reviews*, **1**, 1 (1998), URL http://www.livingreviews.org/Articles; C. Rovelli: *Quantum Gravity*, (Cambridge University Press, Cambridge 2004); R. Gambini and J. Pullin: *Loops, Knots, Gauge Theories and Quantum Gravity*, (Cambridge University Press, Cambridge 1996); C. Beetle and A. Corichi: Bibliography of publications related to classical and quantum gravity in terms of connections and loop variables, gr-qc/9703044.
2. For reviews see for example: L. Dolan: TASI Lectures on perturbative string theory and Ramond-Ramond flux. In: *String, branes and extra dimensions: TASI 2001*, ed by S.S. Gubser and J.D. Lykken (World Scientific, River Edge, N. J. 2004) pp 161-193; R.J. Szabo: BUSSTEPP Lectures on string theory: an introduction to string theory and D-brane dynamics, hep-th/0207142; J. Polchinski: *String Theory, Vols I and II*, (Cambridge University Press, Cambridge 2000).
3. G. Amelino-Camelia: Nature (London) **398**, 216 (1999); D.V. Ahluwalia: Nature (London) **398**, 199 (1999); G. Amelino-Camelia: Lect. Notes Phys. **541**, 1 (2000); G. Amelino-Camelia, Int. Jour. Mod. Phys. **D10**, 1 (2001); N.E. Mavromatos: The quest for quantum gravity: testing times for theories?. In: *From Particles to the Universe*, ed by A. Astbury et al. (World Scientific, Singapore 2001) pp 335-341; J. Ellis: Perspectives in High-Energy Physics. In: *Proceedings of the 3rd Latin American Symposium on High-Energy Physics (SILAFAE III)*, ed by E. Nardi (Institute of Physics, Bristol 2000); G.Z. Adunas, E. Rodriguez-Milla and D.V. Ahluwalia: Phys. Letts. **B485**, 215 (2000); G.Z. Adunas, E. Rodriguez-Milla and

D.V. Ahluwalia: Gen. Rel. Grav. **33**, 183 (2001); J. Ellis: Nuovo Cim. **24C**, 483 (2001); G. Amelino-Camelia: Mod. Phys. Lett. **A17**, 899 (2002); S. Sarkar: Mod. Phys. Lett. **A17**, 1025 (2002); D.V. Ahluwalia: Mod. Phys. Lett. **A17**, 1135 (2002); G. Amelino-Camelia: contribution to these Proceedings.
4. S.D. Biller et al.: Phys. Rev. Lett. **83**, 2108 (1999).
5. O. Betolami and C.S. Carvalho, Phys. Rev. **D61**, 103002 (2000); J.M. Carmona and J.L. Cortés: Phys. Lett. B **494**, 75 (2000); J.M. Carmona and J.L. Cortés: Phys. Rev. **D65**, 025006 (2002); C. Lämmerzahl and C. Bordé: Testing the Dirac equation. In: Lecture Notes in Physics **562**, ed by C. Lämmerzahl, C.W.F. Everitt and F.W. Hehl (Springer, Heidelberg 2001) pp. 463–478; M.P. Haugan and C. Läemmerzhal: Principles of equivalence: their role in gravitation physics and experiments that test them. In: Lecture Notes in Physics **562**, ed by C. Lämmerzahl, C.W.F. Everitt and F.W. Hehl (Springer, Heidelberg 2001) pp. 195–212; G. Lambiase: Gen. Rel. Grav. **33**, 2151 (2001); G. Lambiase: Eur. Phys. J. **C19**, 553 (2001); G. Amelino-Camelia: Phys. Letts. **B528**, 181 (2002); G. Lambiase: Class. Quant. Grav. **20**, 4213 (2003); G. Lambiase and P. Singh: Phys. Letts. **B565**, 27 (2003); G. Amelino-Camelia and C. Lämmerzahl: Class. Quant. Grav. **21**, 899 (2004); F.W. Stecker: Astropart. Phys. **20**, 85 (2003); F.W. Stecker: J. Phys. **G29**, R47 (2003).
6. C. Rovelli and S. Speziale: Phys. Rev. **D67**, 064019 (2003).
7. F. Dowker, J. Henson, and R.D. Sorkin: Mod. Phys. Lett. **A19**, 1829 (2004).
8. V.A. Kostelecký and S. Samuel: Phys. Rev. **D39**, 683 (1989), V.A. Kostelecký and S. Samuel: Phys. Rev. **D40**, 1886 (1989); V.A. Kostelecký and R. Potting: Nucl. Phys. **B359**, 545 (1991); V.A. Kostelecký and R. Potting: Phys. Lett. **B381**, 89 (1996).
9. V.A. Kostelecký and R. Lehnert: Phys. Rev. **D63**, 065008 (2001).
10. V.W. Hughes, H.G. Robinson and V. Beltrán-López: Phys. Rev. Lett. **4**, 342 (1960); R.W.P. Drever: Phil. Mag. **6**, 683 (1961).
11. D. Colladay and V.A. Kostelecký: Phys. Rev. **D55**, 6760 (1997); D. Colladay and V.A. Kostelecký: Phys. Rev. **D58**, 116002 (1998). For reviews see for example *Proceedings of the Meeting on CPT and Lorentz Symmetry*, ed by V.A. Kostelecký (World Scientific, Singapore 1999); C.M. Will, *The Confrontation between General Relativity and Experiment* (Living Reviews in Relativity 4, 2001); *Proceedings of the Second Meeting on CPT and Lorentz Symmetry*, ed by V.A. Kostelecký (World Scientific, Singapore 2002).
12. J.R. Ellis, J.L. Lopez, N.E. Mavromatos and D.V. Nanopoulos: Phys. Rev. **D53**, 3846 (1996).
13. V. A. Kostelecký: Phys. Rev. **D69**, 105009 (2004).
14. R.C. Myers and M. Pospelov: Phys. Rev. Lett. **90**, 211601 (2003); R.C. Myers and M. Pospelov: Experimental Challenges of Quantum Gravity. In: *Quantum Theory and Symmetries*, ed by P.C. Argyres, T.J. Hodges, F. Mansouri, J.J. Scanio, P. Suranyi and L.C.R. Wijewardhana (World Scientific, New Jersey 2004) pp. 732–744; A. Pérez and D. Sudarsky: Phys. Rev. Lett. **91**, 179101 (2003).
15. T. Jacobson, S. Liberati and D. Mattingly: Nature **424**, 1019 (2003).
16. T. Jacobson, S. Liberati and D. Mattingly: Phys. Rev. **D67**, 124011 (2003); T. Jacobson, S. Liberati and D. Mattingly: Phys. Rev. **D66**, 081302 (2002); T. Jacobson, S. Liberati, D. Mattingly and F. Stecker: Phys. Rev. Lett. **93**, 021101 (2004).
17. C. Lämmerzahl, A. Macías and H. Müller: Phys. Rev. **D71**, 025007 (2005) and references therein.

18. A.A. Andrianov and R. Soldati: Phys. Rev. **D51**, 5961 (1995); A.A. Andrianov and R. Soldati: Phys. Letts. **B435**, 449 (1998); A.A. Andrianov, P. Giacconi and R. Soldati: JHEP **0202**, 030 (2002).
19. G. Amelino-Camelia: Int. J. Mod. Phys. **D11**, 35 (2002); S. Judes and M. Visser: Phys. Rev. **D68**, 045001 (2003); A. Agostini, G. Amelino-Camelia and M. Arzano: Class. Quant. Grav. **21**, 2179 (2004); D.V. Ahluwalia-Khalilova: Fermions, bosons and locality in special relativity with two invariant scales. In: *Proceedings of the 4th International Conference on Physics Beyond the Standard Model: Beyond the Desert (BEYOND 03)*, ed by H.V. Klapdor-Kleingrothaus (Springer, Berlin 2004) pp. 503–511; G. Amelino-Camelia, L. Smolin and A. Starodubtsev: Class. Quant. Grav. **21**, 3095 (2004); G. Amelino-Camelia: Three perspectives on the quantum gravity problem and their implications for the fate of Lorentz symmetry, gr-qc/0309054.
20. J. Magueijo and L. Smolin: Phys. Rev. Lett. **88**, 190403 (2002); J. Magueijo and L. Smolin: Phys. Rev. **D67**, 044017 (2003).
21. S. Liberati, S. Sonego and M. Visser: Phys. Rev. **D71**, 045001 (2005).
22. F. Girelli and E. Livine: Physics of deformed special relativity: relativity principle revisited, gr-qc/0412004.
23. L. Smolin: Falsifiable predictions from semiclassical quantum gravity, hep-th/0501091.
24. K. Imilkowska and J. Kowaslki-Glikman: contribution to these Proceedings; J. Kowaslki-Glikman and L. Smolin: Phys. Rev. **D70**, 065020 (2004); J. Kowaslki-Glikman: Introduction to Doubly Special Relativity, hep-th/0405273; J. Kowaslki-Glikman and S. Nowak: Int. J. Mod. Phys. **D12**, 299 (2003).
25. J. F. Donoghue: Phys. Rev. Lett. **72**, 2996 (1994); J. F. Donoghue: Phys. Rev. **D50**, 3874 (1994).
26. D.A.R. Dalvit, F.D. Mazzitelli and C. Molina-Paris: Phys. Rev. **D63**, 084023 (2001).
27. G. Amelino-Camelia, J.R. Ellis, N.E. Mavromatos and D.V. Nanopoulos: Int. J. Mod. Phys. **A12**, 607 (1997); J.R. Ellis, N.E. Mavromatos and D.V. Nanopoulos: Gen. Rel. Grav. **31**, 1257 (1999); J.R. Ellis, N.E. Mavromatos and D.V. Nanopoulos: Phys. Rev. **D62**, 084019 (2000) J.R. Ellis, K. Farakos, N.E. Mavromatos, V.A. Mitsou and D.V. Nanopoulos, Astrophys. J. **535**, 139 (2000); J.R. Ellis, N.E. Mavromatos, D.V. Nanopoulos and G. Volkov: Gen. Rel. Grav. **32**, 1777 (2000); E. Gravanis and N.E. Mavromatos: JHEP **0206**, 019 (2002).
28. J.R. Ellis, N.E. Mavromatos and D.V. Nanopoulos: Gen. Rel. Grav. **32**, 127 (2000); J. Ellis, N.E. Mavromatos and D.V. Nanopoulos: Phys. Rev. **D61**, 027503 (2000).
29. J.R. Ellis, N.E. Mavromatos, D.V. Nanopoulos and A.S. Sakharov: Nature (London) **428**, 386 (2004).
30. R. Gambini and J. Pullin: Phys. Rev. **D59**, 124021 (1999).
31. J. Alfaro, H.A. Morales-Técotl and L.F. Urrutia: Phys. Rev. Lett. **84**, 2318 (2000).
32. J. Alfaro, H.A. Morales-Técotl and L.F. Urrutia: Phys. Rev. **D65**, 103509 (2002).
33. J. Alfaro, H.A. Morales-Técotl and L.F. Urrutia: Phys. Rev. **D66**, 124006 (2002).
34. L.F. Urrutia: Mod. Phys. Lett. **A17**, 943 (2002); L.F. Urrutia: Loop Quantum Gravity Induced Modifications to Particle Dynamics. In: *AIP Conference Proceedings 670A,* ed by U. Cotti, M. Mondragón and G. Tavares-Velasco (American Institute of Physics, New York 2003) pp. 289–297.
35. H. Sahlmann, T. Thiemann and O. Winkler: Nucl. Phys. **B606**, 401 (2001); H. Sahlmann and T. Thiemann: Towards the QFT on curved spacetime limit of

QRG. I: A general scheme, gr-qc/0207030; H. Sahlmann and T. Thiemann: Towards the QFT on curved spacetime limit of QRG.II: A concrete implementation, gr-qc/0207031.
36. T. Padmanabhan: Phys. Rev. **D57**, 6206 (1998); K. Srinivasan, L. Sriramkumar and T. Padmanabhan: Phys. Rev. **D58**, 044009 (1998); S. Shankaranarayanan and T. Padmanabhan: Int. J. Mod. Phys. **D10**, 351 (2001).
37. C. Chryssomalakos and E. Okon: Int. J. Mod. Phys **D13**, 2003 (2004); A. Corichi and D. Sudarsky: New quantum gravity phenomenology, gr-qc/0503078.
38. P. Huet and M. Peskin: Nucl Phys. **B434**, 3 (1995); J.R. Ellis, J. López, N.E. Mavromatos and D.V. Nanopoulos: Phys. Rev. **D53**, 3846 (1996).
39. S. Coleman and S.L. Glashow, Phys. Rev. **D59**, 116008 (1999).
40. G. Amelino-Camelia, J.R. Ellis, N.E. Mavromatos, D.V. Nanopoulos and S. Sarkar: Nature (London) **393**, 763 (1998).
41. J.R. Ellis, N.E. Mavromatos, D.V. Nanopoulos and A.S. Sakharov: Astron. Astrophys. **402**, 409 (2002); J.R. Ellis, N.E. Mavromatos and A.S. Sakarov: Astropart. Phys. **20**, 669 (2004).
42. S.E. Boggs, C.B. Wunderer, K. Hurley and W. Coburn: Astrophys. J. **611**, L77 (2004)
43. E. Waxman and J. Bahcall: Phys. Rev. Lett. **78**, 2292 (1997); E. Waxman: Nucl. Phys. (Proc. Supl) **91**, 494 (2000); E. Waxman: Nucl. Phys. (Proc. Supl) **87**, 345 (2000).
44. M. Vietri: Phys. Rev. Letts. **80**, 3690 (1998).
45. R. Bluhm: contribution to these Proceedings; R. Bluhm: QED Test of Lorentz Symmetry, hep-ph/041149; R. Bluhm, V.A. Kostelecký, C.D. Lane and N. Russell, Phys. Rev. Lett. **88**, 090801 (2002); R. Bluhm and V.A. Kostelecký, Phys. Rev. Lett. **84**, 1381 (2000).
46. R.J. Gleiser and C.N. Kozameh: Phys. Rev. **D64**, 083007 (2001).
47. T. Kifune: Astrophys. J. Lett. **518**, L21 (1999).
48. G. Amelino-Camelia and T. Piran: Phys. Rev. **D64**, 036005 (2001).
49. G. Amelino-Camelia: Phys. Lett. **B 528**, 181 (2002).
50. J. Alfaro and G. Palma: Phys. Rev. **D65**, 103516 (2002); J. Alfaro and G. Palma: Phys. Rev. **D67**, 083003 (2003).
51. T.J. Konopka and S.A. Major: New J. Phys. **4**, 57 (2002).
52. R. Montemayor and L.F. Urrutia: Phys. Lett. B. **606**, 86 (2005); R. Montemayor and L.F. Urrutia: Radiation in Lorentz Violating Electrodynamics. In: *Gravitation and Cosmology, AIP Conference Proceedings 758*, ed by A. Macías, C. Lämmerzhal and D. Nuñez (American Institute of Physics, New York, 2005) pp. 81–89; R. Montemayor and L.F. Urrutia: Synchrotron radiation in Lorentz-violating Electrodynamics: the Myers–Pospelov model, hep-ph/0505135.
53. B. González, S. A. Martínez, R. Montemayor and L.F. Urrutia: Lorentz Violating Electrodynamics, hep-ph/0505145.
54. H. Vucetich: Testing Lorentz invariance violation in quantum gravity theories, gr-qc/0502093; D. Mattingly: Modern test of Lorentz invariance, gr-qc/0502097; T. Jacobson, S. Liberati and D. Mattingly: Lorentz violation at high energy: concepts, phenomena and astrophysical constraints, astro-ph/0505277.
55. T. Thiemann: Introduction to Modern Canonical Quantum General Relativity, gr-qc/0110034; A. Pérez: Introduction to Loop Quantm Gravity and Spin Foams, grqc/0409061.
56. C. Rovelli: Helv. Phys. Acta **69**, 582 (1996); C. Rovelli: Phys. Rev. Lett. **77**, 3288 (1996); A. Ashtekar, J. Baez, A. Corichi and K. Krasnov: Phys. Rev. Lett. **80**, 904 (1998).

57. M. Bojowald: Phys. Rev. Lett. **86**, 5227 (2001); For a recent review see for example M. Bojowald and H.A. Morales-Técotl: Cosmological applications of loop quantum gravity. In: *Proceedings of 5th Mexican School on Gravitation and Mathematical Physics: The Early Universe and Observational Cosmology (DGFM 2002)*, Lect. Notes Phys. **646**, ed by N. Bretón, J.L. Cervantes-Cota and M. Salgado (Springer, Berlin 2004) pp. 421–462.
58. C. Rovelli and L. Smolin: Nucl. Phys. **B442**, 593 (1995); Erratum-ibid **B456**, 753 (1995); A. Ashtekar and J. Lewandowsky: Class. Quant. Grav. **14**, A55 (1997); A. Ashtekar and J. Lewandowsky: Adv. Theor. Math. Phys. **1**, 388 (1998); A. Ashtekar, A. Corichi and J. Zapata: Class. quant. Grav. **15**, 2955 (1998).
59. A. Ashtekar: Phys. Rev. Letts. **57**, 2244 (1987); F. Barbero: Phys. Rev. **D51**, 5507 (1995); F. Barbero: Phys. Rev. **D51**, 5498 (1995).
60. G. Immirzi: Class. Quant. Grav. **14**, L177 (1997).
61. T. Thiemann: Phys. Lett. **B380**, 257 (1996).
62. T. Thiemann: Class. Quan. Grav. **15**, 1281 (1998); T. Thiemann: Class. Quant. Grav. **15**, 839 (1998).
63. M. Varadarajan and J.A. Zapata: Class. Quant. Grav. **17**, 4085 (2000); M. Varadarajan: Phys. Rev. **D61**, 104001 (2000); T. Thiemann: Class. Quant. Grav. **18**, 2025 (2001); T. Thiemann and O. Winkler: Class. Quant. Grav. **18**, 2561 (2001); T. Thiemann and O. Winkler: Class. Quant. Grav. **18**, 4629 (2001); T. Thiemann and O. Winkler: Class. Quant. Grav. **18**, 4997 (2001); H. Sahlmann, T. Thiemann, O. Winkler: JHEP **0105**, 021 (2001); A. Ashtekar and J. Lewandowski: Class. Quant. Grav. **18**, L117 (2001); A. Ashtekar, S. Fairhurst and J.L. Willis: Class. Quant. Grav. **20**, 1031 (2003); A. Ashtekar, L. Bombelli and A. Corichi: Semiclassical states for constrained systems, gr-qc/0504052.
64. J. Alfaro, H.A. Morales-Técotl, M. Reyes and L.F. Urrutia: J. Phys. A: Math. Gen. **36**, 12097 (2003).
65. G. Lambiase: Mod. Phys. Lett. **A18**, 23 (2003).
66. G.L. Fogli, E. Lisi, A. Marrone and G. Scioscia: Phys. Rev. **D60**, 053006 (1999); R. Brustein, D. Eichler and S. Foffa: Phys. Rev. **D65**, 105006 (2002).
67. C.H. Lineweaver et al.: Astrophys J. **470**, 38 (1996).
68. D. Sudarsky, L. Urrutia and H. Vucetich: Phys. Rev. Lett. **89**, 231301 (2002).
69. J.D. Prestage et al.: Phys. Rev. Letts. **54**, 2387 (1985); S. K. Lamoreaux et al.: Phys. Rev. Lett. **57**, 3125 (1986); S. K. Lamoreaux et al.: Phys. Rev. **A 39**, 1082 (1989); C. J. Berglund et al.: Phys. Rev. Lett. **75** 1879 (1995); D. F. Phillips et al.: Phys. Rev D **63** 111101 (2001); R. Walsworth: contribution to this volume, p. 493.
70. T.E. Chupp et al.: Phys. Rev. Lett. **63**, 1541 (1989).
71. D. Bear et al.: Phys. Rev. Lett. **85** 5038 (2000).
72. V.A. Kostelecký and C.D. Lane: J. Math. Phys. **40**, 6245 (1999); V.A. Kostelecký and R. Lehnert: Phys. Rev. **D63**, 065008 (2001); D. Colladay and P. McDonald: J. Math. Phys. **43**, 3554 (2002); For a review see for example V.A. Kostelecký: Topics in Lorentz and CPT violations. In: *The Role of Neutrinos, Strings, Gravity and Variable Cosmological Constant in Elementary Particle Physics*, ed by B.N. Kursunoglu, S.L. Mintz and A. Perlmutter (Kluwer, New York 2001) pp. 57–68.
73. D. Sudarsky, L. Urrutia and H. Vucetich: Phys. Rev. **D68**, 024010 (2003).
74. J. Collins, A. Perez, D. Sudarsky, L. Urrutia and H. Vucetich: Phys. Rev. Lett. **93**, 191301 (2004).
75. J. Collins, A. Perez, D. Sudarsky, L. Urrutia and H. Vucetich: unpublished.
76. S.G. Nibbelink and M. Pospelov: Phys. Rev. Letts. **94**, 081601 (2005); P. Jain and J.P. Ralston: Supersymmetry and the Lorentz fine tuning problem, hep-

th/0502106; P.A. Bolokov, S.G. Nibbelink and M. Pospelov: Lorentz violating supersymmetric quantum electrodynamics, hep-ph/0505029.
77. J. Alfaro: Phys. Rev. Lett **94**, 221302 (2005); J. Alfaro: LIV dimensional regularization and quantum gravity effects in the standard model, hep-th/0501129; J. Alfaro: Quantum gravity induced Lorentz invariance violation in the Standard Model: hadrons, hep-th/0505228.
78. G. Amelino-Camelia: New J. Phys. **6**, 188 (2004); T. Jacobson, S. Liberati and D. Mattingly: Comments on "Improved limit on quantum-spacetime modifications of Lorentz symmetry from observations on gamma-ray blazers", gr-qc/0212002 v2.
79. W. Coburn and S.E. Boggs: Nature (London) **423**, 415 (2003); W. Coburn and S.E. Boggs: Statistical uncertainty in the re-analysis of polarization in GRB021206, astro-ph/0310515.
80. R.E. Rutledge and D.B. Fox: Mon. Not. Roy. Astron. Soc. **350**, 1272 (2004); C. Wigger et al.: Astrophys. J. **613**, 1088 (2004).
81. J. Schwinger: Phys. Rev. **75**, 1912 (1949); J. Schwinger: On the radiation by electrons in a betatron, transcribed by M. A. Furman. In: *A Quantum Legacy*, ed by K. Milton, (World Scientific, Singapore, 2000).
82. J. Schwinger, L.L. DeRaad, Jr., K.A. Milton and Wu-yang Tsai: *Classical Electrodynamics*, (Westview Press, Boulder, 1998).
83. J. Schwinger, W-Y. Tsai and T. Erber: Ann. Phys. (NY) **96**, 303 (1976).
84. See for example T. Erber: Rev. Mod. Phys. **38**, 626 (1966).
85. T. Erber, D. White and H.G. Latal: Acta Phys. Austr. **45**, 29 (1976).
86. S.M. Carroll, G.B. Field and R. Jackiw: Phys. Rev. **D41**, 1231 (1990); V.A. Kostelecký and M. Mewes, Phys. Rev. Lett. **87**, 251304 (2001); R. Lehnert and R. Potting: Phys. Rev. Lett. **93**, 110402 (2004); Q.G. Bailey, and V.A. Kostelecky: Phys. Rev. **D70**, 076006 (2004).
87. A.M. Atoyan and F.A. Aharonian: Mon. Not. R. Astron. Soc. **278**, 525 (1996); P.L. Marsden, F.C. Gillett, F.E. Jennings, J.P. Emerson, T. de Jong and F.N. Olnon: Astrophys. J. **278**, L29 (1984).
88. F.A. Aharonian: New Astron. **5**, 377 (2000).
89. A. Konopelko, A. Mastichiadis, J. Kirk, O.C. de Jager and F.W. Stecker: Astrophys. J. **597**, 851 (2003).
90. E. Nakar, T. Piran and E. Waxman: JCAP **0310**, 005 (2003).
91. E. Waxman: Nature (London) **423**, 388 (2003).
92. W. Hajdas et al.: Time resolved annalysis of the Gamma Ray Bursts with the RHESSI satellite. In: *Particle acceleration in astrophysical objects*, (http://www.oa.uj.edu.pl/konferencje/proc0.html).
93. P. Castorina, A. Iorio and D. Zappala: Phys. Rev. **D69**, 065008 (2004).
94. G.B. Rybicki and A.P. Lightman: *Radiative Processes in Astrophysics*, (John Wiley and Sons, New York, 1979); A.L. Peratt: *Physics of the Plasma Universe*, (Springer Verlag, New York 1991)
95. S.A. Martínez, R. Montemayor and L.F. Urrutia: Reality and causality in LIV radiation, in preparation.

Part IV

Experimental Search

Test Theories for Lorentz Invariance

C. Lämmerzahl

ZARM, University of Bremen, Am Fallturm, 28359 Bremen, Germany
laemmerzahl@zarm.uni-bremen.de

Abstract. After a very short review of the principles underlying Special Relativity, their meaning, and their consequences, we first describe the basic experiments testing SR in a model–independent way which is the most basic way to describe experiments testing the foundations of SR. In order to be able to give quantitative estimates of the validation of SR and, even more important, in order to be able to compare conceptually different experiments, one introduces test theories. We give a review of test theories needed for a consistent description of tests of Lorentz Invariance. The main emphasize is on kinematical test theories of Robertson and Mansouri–Sexl type. Though these test theories were very important in reaching a new understanding of the experimental foundation of SR, an extensive discussion shows that are kinematical test theories are incomplete and, thus, dynamical test theories like the Standard Model Extension are superior.

1 Introduction

1.1 Postulates of Special Relativity

Special Relativity (SR) is one of the rare examples where essentially everything, the formalism as well as all physical consequences, can be bases on two postulates only. These two postulates are

Postulate 1: The speed of light c is constant.
Postulate 2: The relativity principle.

The first postulate may be replaced by a perhaps even more simple one, namely by the statement that light is a unique phenomenon, that is, between an event and a worldline there are two and only two light rays. The light ray, in particular, does not depend on the trajectory the event of the emission point lies on. Otherwise there will be more than two light rays. The second postulate then makes sure that the measured velocity of light does not depend on its direction and on the velocity of the observer.

The two postulates have some immediate consequences which all can be tested in experiments:

C. Lämmerzahl: *Test Theories for Lorentz Invariance*, Lect. Notes Phys. **702**, 349–384 (2006)
DOI 10.1007/3-540-34523-X_12 © Springer-Verlag Berlin Heidelberg 2006

- The velocity of light, c, does not depend on
  - the velocity of the source (what is a statement of the uniqueness of the phenomenon)
  - the velocity of the observer,
  - the direction of propagation,
  - the polarization or frequency of the light ray.
- The relativity principle implies that
  - the limiting velocity of all particles is the speed of light

$$c = c_+ = c_- = c_\nu = v_p^{\max} = v_e^{\max} = v_{\text{grav}}$$

    (otherwise there is a preferred frame in contradiction to the second postulate), with the consequence that
  - c is universal and, thus, can be interpreted as *geometry*,
  - that *all* physics is the same in *all* inertial systems, that is, experimental results do not depend on the
    - orientation of the laboratory and
    - on the velocity of the laboratory.

The experimental status of the foundations of SR has been reviewed recently in [1–3] and a description of technological applications of SR can be found in [4].

1.2 The Consequences

From the above postulates one can derive the Lorentz transformations

$$t' = \frac{1}{\sqrt{1-v^2}}\,(t - \boldsymbol{x}\cdot\boldsymbol{v}) \tag{1}$$

$$\boldsymbol{x}' = \boldsymbol{x}_\perp + \frac{1}{\sqrt{1-v^2}}\left(\boldsymbol{v}_\parallel - \boldsymbol{v}t\right), \tag{2}$$

where $\boldsymbol{x}_\parallel = \boldsymbol{x}\cdot\boldsymbol{v}/v^2$ and $\boldsymbol{x}_\perp = \boldsymbol{x} - \boldsymbol{x}_\parallel$. These transformations lead to the following effects

1. time dilation,
2. twin paradox,
3. Doppler effect,
4. length contraction,
5. addition of velocities,
6. Sagnac effect, and
7. Thomas precession.

All these effects except length contraction have been confirmed in experiments with high accuracy.

1.3 The Ether

One consequence of the Galilei–transformations is the addition of velocities: If a body moves with velocity $\boldsymbol{u}$ with respect to an inertial system S, then another inertial system S' moving with $-\boldsymbol{v}$ with respect to S observes the body with a velocity

$$\boldsymbol{u}' = \boldsymbol{u} + \boldsymbol{v}\,. \tag{3}$$

This applies to all velocities, in particular to the speed of light. That frame in which the speed of light is isotropic and is what appears in Maxwell's equations, is called the ether frame.

If $\boldsymbol{c}$ is the speed of light in the ether frame, then in a frame moving with respect to the ether the velocity of light is $\boldsymbol{c}' = \boldsymbol{c} + \boldsymbol{v}$ with an orientation and velocity–dependent modulus

$$\begin{aligned} c'(\theta, v) &= \sqrt{c^2 + v^2 + 2cv\cos\theta} \\ &\approx c\left(1 + \frac{v}{c}\cos\theta + \frac{1}{2}\frac{v^2}{c^2}\left(1 + 3\cos\theta\right)\right) + \mathcal{O}(v^4/c^4)\,, \end{aligned} \tag{4}$$

where $\theta = \measuredangle(\boldsymbol{c}, \boldsymbol{v})$. This orientation dependence was looked for in the Michelson–Morley experiments.

2 Test Theories

2.1 What are Test Theories?

Test theories are parametrized generalizations or "violations" of theories under consideration. Calculations of experiments using these generalized theories lead to a variety of effects which are absent in the ordinary theory. A comparison of the calculated effects with experimental results leads to estimates of the parameters characterizing the violation of the theory. One main aspect is that only one generalized theory is taken in order to describe all possible effects.

Consequently, tests theories have the following advantages and tasks

1. Parametrization and identification of possible violation.
2. Quantification of degree of validity.
3. Different (!) experiments can be compared.

In particular the last point is important since in principle different tests may need different theories. For example, while the Michelson–Morley experiment examines the outcome of an interference experiment during a change of the orientation of the apparatus, the Kennedy–Thorndike experiments examines the same for a change in the velocity of the apparatus. Both situations are different and have nothing to do with one another. Only within tests theories both possible results can be described by a (different) combination of one set of parameters.

Different test theories contain a different number of parameters characterizing the deviation from the standard theory. Accordingly, for different test theories

one needs a different number of independent tests in order to verify within the experimental limits the theory under consideration. Each test theory defines itself the experiments needed for that.

The quality or richness of a test theory depends on the number of parameters. More parameters can describe a wider range of hypothetical effects and therefore a more complete characterization or verification of the theory under consideration is possible. However, in some cases it is preferable to restrict to a small set of parameters in order to focus on distinguished features. Examples of this are the Robertson test theory or the c^2 formalism [5, 6] which is a special case of the $TH\epsilon\mu$–formalism [7] and which is equivalent a one–parameter subset of the Standard Model Extension, see Table 1. The fully parametrized Standard Model Extension is rather cumbersome to treat. It is obvious that one needs as many independent tests as there are parameters which have to be determined. In the c^2 formalism only one test is needed; for the RMS theory we need three tests and for the SME one needs at the end more than 100 tests. One of the theoretical tasks is to find out that tests which may yield the best estimate for the parameters under consideration.

Beside the reasons mentioned above, test theories also play the role to mediate between the experimental results and a full theory of, e.g., quantum gravity, see Fig. 1.

There are two classes of test theories for SR, a kinematical and a dynamical. Kinematical test theories have been worked out by Robertson [8] and by Mansouri and Sexl [9], dynamical test theories are the $TH\epsilon\mu$–Formalism [7, 10], the Extended Standard Model [11] or even more general setups [12].

2.2 Kinematical Test Theories

Kinematical tests theories discuss the transformation between inertial frames moving with different velocities. At first, these transformation possess the general structure $x'^a = f^a{}_b(v)x^b$, where v is the relative velocity between the two frames. Each kinematical test theory considers a certain class of these general transformations. These transformations are used in order to describe experiments in different frames which, in general, may depend on v. The kinematical test theory of Robertson and Mansouri and Sexl were a very important step for the understanding of the structure of SR. Within this test theory the three famous classical experiments of Michelson–Morley, Kennedy–Thorndike, and Ives

Table 1. Test theories and their number of parameters. In the SME n = number of different elementary particles like electrons, protons, neutrons, etc.

Test Theory	Number of Parameters
c^2–formalism	1 parameter
Robertson–Mansouri–Sexl	3 parameters
$\chi - g$–formalism	19 parameter
Extended Standard Model	$19 + n48$ parameter

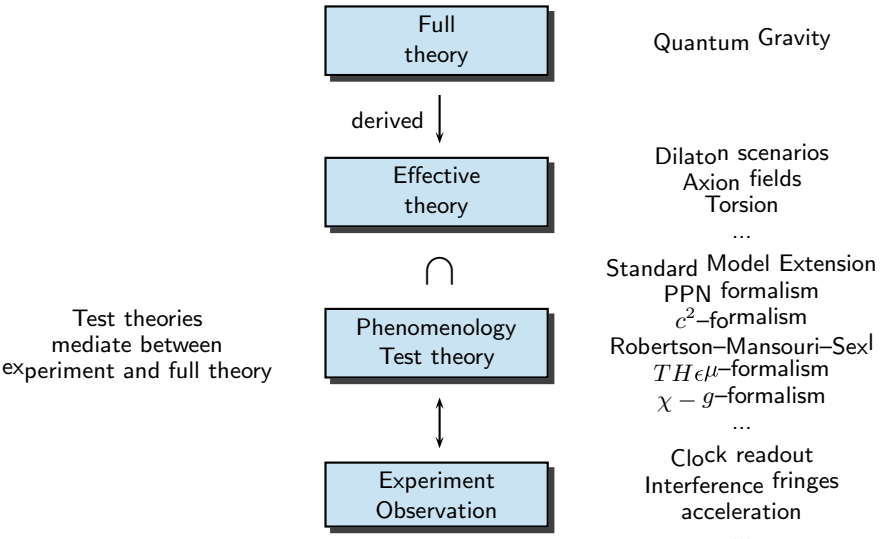


Fig. 1. The hierarchy of descriptions of the physical world. The effective theories are a subset of the phenomenological theories

and Stilwell are identified which are needed in order to verify SR. For dynamical test theories this classification turns out to be not sufficient, in particular since only the behavior of light is considered.

2.3 Dynamical Test Theories

Dynamical test theories start with generalized equations of motion which are used in order to describe experiments. This means that generalized equations of motion for the present standard model are needed, in particular generalized Maxwell and Dirac equations. There are of course infinitely many ways to generalize equations. However, for each kind of phenomenon related to the violation of one of the principles underlying SR, one can begin with very simple modifications. Starting from the standard Maxwell and Dirac equations these modifications in the following (5) and (5) may consist of introducing

- terms $\chi^{\mu\rho\nu\sigma}$, M, and X^{ab} violating Lorentz invariance (see the contribution of R. Bluhm in this volume),
- terms $\chi^{\mu\rho\sigma}$ violating charge non–conservation [12] which also violate Lorentz invariance,
- higher derivatives which in general also violate Lorentz invariance,
- non–linearities.

These modifications then yield the generalized Maxwell and Dirac equations

$$4\pi j^{\mu} = \eta^{\mu\rho}\eta^{\nu\sigma}\partial_{\nu}F_{\rho\sigma} + \chi^{\mu\rho\nu\sigma}\partial_{\nu}F_{\rho\sigma} + \chi^{\mu\rho\sigma}F_{\rho\sigma} + \chi^{\mu\rho\nu\sigma\tau}\partial_{\nu}\partial_{\tau}F_{\rho\sigma} + \ldots + \zeta^{\mu\rho\sigma\tau\nu}F_{\rho\sigma}F_{\tau\nu} + \ldots \quad (5)$$

$$0 = i\gamma^{a}D_{a}\psi + m\psi + M\psi + \gamma^{ab}D_{a}D_{b}\psi + \ldots + N(\psi)\psi \quad (6)$$

where $D_a = \partial_a - ieA_a$ and

$$\gamma^{a}\gamma^{b} + \gamma^{b}\gamma^{a} = 2\eta^{ab} + X^{ab}\,. \quad (7)$$

The possible effects which can be derived from the above generalized equations are

- Birefringence
- Anisotropic speed of light
- Anisotropy in quantum fields
- Charge non-conservation
- Anomalous dispersion
- Decoherence, space-time fluctuations
- Modified interference
- Non-localities

In general, as in the Standard Model Extension, for example, the parameters are assumed to be constant.

In this contribution we in extenso treat the kinematical test theories, make some remarks on dynamical test theories and, finally, present a comparison between these test theories.

3 Model-Independent Descriptions of LI Tests

Before we enter the description of the tests of LI in terms of kinematical tests theories, we describe them in a model independent way. Here "model independent" means that we do not assume anything related to the space-time geometry. We of course employ models related to wave propagation, resonators, etc. which – and we like to emphasize this once more – do not anticipate any results on the Lorentzian structure of space–time.

3.1 Isotropy of the Speed of Light

There are two main experimental schemes for testing the isotropy of the speed of light: rotating Michelson interferometers and rotating resonators. We describe both.

Interference Experiments

The Setup

The setup of the experiment by Michelson and Morley [13] uses a Michelson interferometer mounted on a turn table, see Fig. 2. Light from a source is split

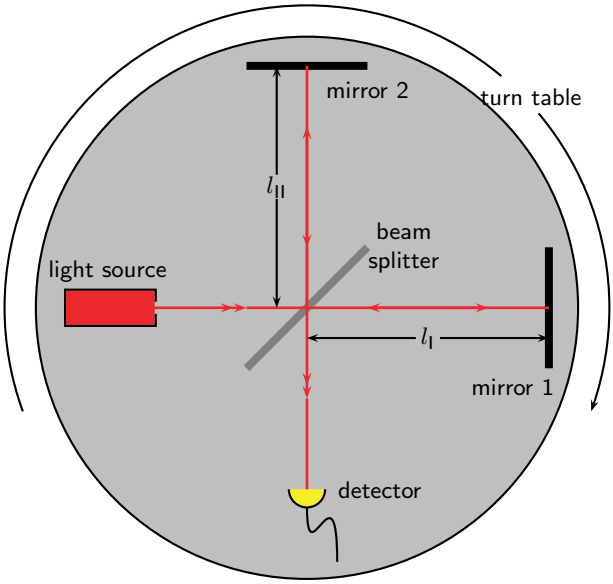


Fig. 2. Setup of the Michelson–Morley experiment. A Michelson interferometer is mounted on a turn table. One looks for a variation of the intensity for varying orientation

coherently and propagates along two different directions. After reflection by mirrors the light rays recombine and interfere. The intensity of the interfering light rays is observed in the detector.

Model Independent Description

We assume the interferometer is in the $x-y$–plane. The incoming light ray is described by a plane wave[1] with frequency ω $e^{i(kx-\omega t)}$. The two split light waves are given by $e^{i(k_{1,2\pm}x-\omega t)}$, where k_{1+} is the wave vector of the wave propagating from the beam splitter to the mirrors, and $k_{1,2-}$ is the wave vector of the reflected waves. Stationarity requires a unique frequency.

The intensity of the interfering waves is

$$I = \frac{1}{2}\left|e^{i(k_{1+}l_1+k_{1-}l_1+\omega t)} + e^{i(k_{2+}l_2+k_{2-}l_2+\omega t)}\right|^2 , \tag{8}$$

where l_1 and l_2 are the lengths of the interferometer arms. We use the dispersion relations[2] $\omega = k_{1\pm}\, c_{1\pm}$ and $\omega = k_{2\pm}\, c_{2\pm}$, where c_{1+} and c_{2+} are the velocities

[1] Here we assume that light can be described by a plane wave. This assumption is independent from any results concerning Lorentz invariance.

[2] Dispersion relations are a consequence of dynamical equations like the wave equation if one discusses plane wave solutions.

of light (phase velocity) propagating from the beam splitter to the mirrors and c_{1-} and c_{2-} the corresponding velocities in opposite direction, and obtain

$$I = \frac{1}{2}\left| e^{i\omega\left(\frac{l_1}{c_{1+}}+\frac{l_1}{c_{1-}}-t\right)} + e^{i\omega\left(\frac{l_2}{c_{2+}}+\frac{l_2}{c_{2-}}-t\right)}\right|^2 = \frac{1}{2}\left[1+\cos\left(\frac{2\omega l_2}{c_2}-\frac{2\omega l_1}{c_1}\right)\right]. \tag{9}$$

Here c_1 and c_2 are the synchronization independent two-way velocities

$$\frac{2}{c_{1,2}} = \frac{1}{c_{1,2+}} + \frac{1}{c_{1,2-}} \tag{10}$$

along the two interferometer arms. The observable phase shift is

$$\Delta\phi = \omega\left(\frac{2l_2}{c_2}-\frac{2l_1}{c_1}\right), \tag{11}$$

We assume a small variation of the speed of light, $c_{1,2} = c + \delta c_{1,2}$ with $\delta c_{1,2} \ll c$ and obtain

$$\Delta\phi = 2\omega\left(\frac{l_2-l_1}{c} + \frac{l_1}{c}\frac{\delta c_1}{c} - \frac{l_2}{c}\frac{\delta c_2}{c}\right). \tag{12}$$

The variation $\delta c_{1,2}$ may depend on the orientation θ of the interferometer. Since the interferometer arms are orthogonal, $\delta c_2 = \delta c(\theta)$ and $\delta c_1 = \delta c(\theta+\frac{\pi}{2})$. Then

$$\Delta\phi(\theta) = 2\omega\left(\frac{l_2-l_1}{c} + \frac{l_1}{c}\frac{\delta c(\theta+\frac{\pi}{2})}{c} - \frac{l_2}{c}\frac{\delta c(\theta)}{c}\right). \tag{13}$$

Upon rotating the interferometer an orientation dependent speed of light yields the phase shift

$$\delta\Delta\phi(\theta) = 2\omega\left(\frac{l_1}{c}\frac{\delta c(\theta+\frac{\pi}{2})}{c} - \frac{l_2}{c}\frac{\delta c(\theta)}{c}\right) \overset{l_2=l_2}{=} 2\frac{\omega l}{c}\left(\frac{\delta c(\theta+\frac{\pi}{2})}{c} - \frac{\delta c(\theta)}{c}\right), \tag{14}$$

where in the last step we also assumed an equal arm interferometer.

In the derivation we assumed that the speed of light might depend on the orientation. It can already be seen from (11) that an orientation dependent arm length give the same effect. Operationally one cannot distinguish between a variation of the speed of light and the a variation of the arm length. What can be observed is the difference in the changes. In fact, in dynamical approaches both has to be taken into account [14]. This also means that in experiments the length of the interferometer has to be controlled very carefully. Any thermal change of the length may simulate a varying speed of light. It is a general agreement to formally assigning any result to the speed of light, that is, we *define* the length of the interferometer arm as constant provided any external influence has been ruled out (what sometimes is subject to some debates as, e.g., in the case of the Miller experiment [15]).

In the case $\delta c(\theta) = \delta c \cos\theta$ we obtain

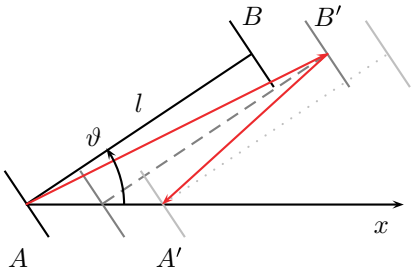


Fig. 3. The paths of light in an interferometer arm moving to the right in an ether frame. The angle between the velocity of the interferometer with respect to the ether and the orientation of the interferometer arm is ϑ

$$\delta\phi(\theta) = -2\omega\frac{l}{c}\left(\sin\theta + \cos\theta\right)\frac{\delta c}{c} = \frac{2l\omega}{c}\sqrt{2}\sin(\theta + \tfrac{\pi}{4})\frac{\delta c}{c}\,. \tag{15}$$

This is the expected phase shift for $\delta c \neq 0$. If no phase shift is observed, then $\delta c = 0$ within the accuracy of the interferometer. (For an interferometer with orthogonal arms a variation $\delta c(\theta) = \delta c\cos(4n\theta)$, $n \in \mathbb{N}$, cannot be detected.)

Interpretation within the Ether Theory

In the ether frame the calculation of the time $t(\vartheta, v)$ light needs to propagate from the beam splitter to one mirror and back to the beam splitter immediately yields from Fig. 3

$$t(l,\vartheta) = \frac{2lc}{c^2 - v^2}\sqrt{1 - \frac{v^2}{c^2}\left(1 - \cos^2\vartheta\right)}\,. \tag{16}$$

where ϑ is the angle between the interferometer arm and $\boldsymbol{v}$. The difference of the time for light moving along two orthogonal interferometer arms is $\Delta t = t(l,\vartheta) - t(l,\vartheta + \pi/2)$. This gives the phase shift

$$\Delta\phi = \frac{2l\omega}{c}\frac{1}{1 - \frac{v^2}{c^2}}\left(\sqrt{1 - \frac{v^2}{c^2}\left(1 - \cos^2\vartheta\right)} - \sqrt{1 - \frac{v^2}{c^2}\left(1 - \sin^2\vartheta\right)}\right) \tag{17}$$

$$= \frac{2l\omega}{c}\frac{v^2}{c^2}\left(\cos^2\vartheta - \sin^2\vartheta\right) = \frac{2l\omega}{c}\frac{v^2}{c^2}\cos(2\vartheta)\,. \tag{18}$$

The same result comes out when we perform the calculation in the frame of the interferometer and make use of the speed of light given by (4).

For an interferometer with an arm length of 11 m as used by Michelson and Morley, and a wavelength of 550 nm one obtains a phase shift of $\Delta\phi = 0.8\pi$ if one uses the velocity of approx. 30 km/s of the Earth around the Sun. Today one would have taken the velocity of the Earth of approx. 360 km/s with respect to the cosmological background which is one order of magnitude larger and, thus,

yields a phase shift of approx 10π. The sensitivity of the original Michelson–Morley interferometer was $\Delta\phi \sim 0.01\pi$ so that this effect should be measurable. However, nothing has been seen which means that $v \leq 8$ km/s.

This null result has been explained by a drag of the ether. Another hypothesis was the length contraction suggested by Lorentz and FitzGerald. Since this contraction should be universal, experiments have been carried through with different materials for the interferometer arms [16,17].

A comparison of the phase shifts gives a relation of velocity of the motion of the reference frame with respect a hypothetical ether to the orientation dependent variation of the speed of light

$$\frac{v^2}{c^2} = \sqrt{2}\frac{\delta_\theta c}{c}\,. \tag{19}$$

Experiments with Resonators

The Setup

In 1955 Essen for the first time used (microwave–) resonators instead of interferometers in order to search for an anisotropic speed of light [18], see Fig. 4. The frequency of a standing electromagnetic wave inside the resonator is determined by the length of the resonator and the speed of light. This frequency can be measured. A varying frequency during turning around the resonator signals an

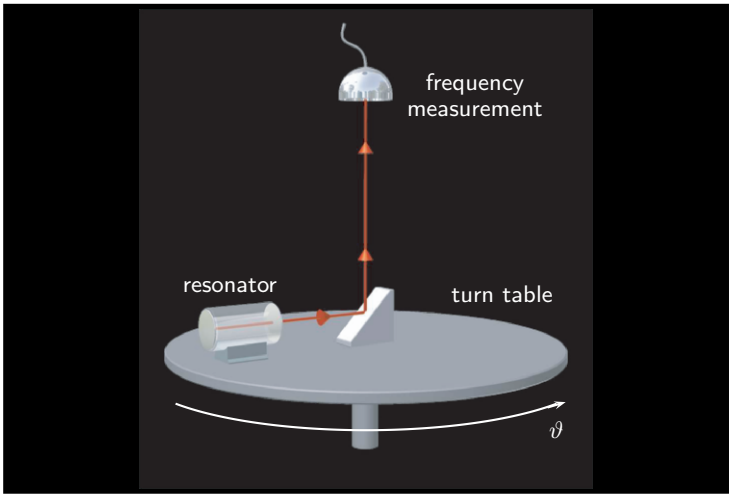


Fig. 4. The principal setup for a test of the isotropy of light using resonators. The frequency of electromagnetic radiation inside the resonators is given by the ratio of the speed of light and the length of the resonator. A change of the frequency during a change of the resonators implies that either the speed of light or the length of the resonator changes with orientation

orientation dependent speed of light (or an orientation dependent length of the resonator). In a modified setup one can use two orthogonally oriented resonators what resembles the Michelson–Morley setup.

Why resonators? Experiments with resonators are much more precise than experiments using interferometers. There are two main reasons: (i) The high finesse (quality factor) of the resonator which today is of the order 10^5 to 10^6. This means that a photon can travel 10^5 to 10^6 times back and forth between the mirrors before leaving the resonator. Therefore, the effective optical path length is much longer than in interferometers. Therefore, a photon can accumulate much more information on an anisotropic speed ob light than in interferometers. For a resonator of 10 cm this amounts to 10 to 100 km compared to 10 m arm length of a typical interferometer. (ii) Resonators are much smaller than interferometers so that much better temperature, vibration, etc. control can be applied. These two reasons lead to the present accuracy of these devices. As an illustration: the distance between the two mirrors can be controlled to up to 1/100 of a proton radius.

Model Independent Description

We have to determine the frequency of the standing electromagnetic wave inside the resonator. This wave consists of two parts traveling back and forth

$$\varphi = Ae^{-i(\omega_+ t - k_+ x)} + Be^{-i(\omega_- t + k_- x)} \,. \tag{20}$$

For a stationary problem we have $\omega = \omega_+ = \omega_-$. Again we use the dispersion relation $\omega = k_\pm c_\pm$ (see footnote on page 355). The velocities of light $c_\pm$ may depend on the orientation related to the orientation of the resonator. Then

$$\varphi = Ae^{-i\omega\left(t - \frac{x}{c_+}\right)} + Be^{-i\omega\left(t + \frac{x}{c_-}\right)} \,. \tag{21}$$

The amplitudes A and B have to be determined using the ordinary boundary conditions $\varphi(0) = 0$ and $\varphi(L) = 0$. The first condition yields $B = -A$ so that

$$\varphi = Ae^{-i\omega t}\left(e^{i\omega\frac{x}{c_+}} - e^{-i\omega\frac{x}{c_-}}\right) \,. \tag{22}$$

The boundary condition at $x = L$

$$0 = e^{i\omega\frac{L}{c_+}} - e^{-i\omega\frac{L}{c_-}} \tag{23}$$

is fulfilled if

$$\sin\left(\omega\left(\frac{1}{c_+} + \frac{1}{c_-}\right)\frac{L}{2}\right) = 0 \,, \tag{24}$$

or, equivalently,

$$\frac{\omega}{c} = \frac{n\pi}{L} \,, \qquad n \in \mathbb{N} \tag{25}$$

with the two-way velocity c. This corresponds to a frequency

$$\nu(\theta) = \frac{n}{2L}c(\theta) \,, \tag{26}$$

where we assumed an orientation dependent two-way speed of light. While turning the resonator on a turn table the frequency of the outcoupled electromagnetic wave is compared with a stationary mounted frequency standard.

In the case of two orthogonally oriented resonators one can observe the beat frequency

$$\nu(\theta+\tfrac{\pi}{2})-\nu(\theta)=\frac{n}{2L}\left(c(\theta+\tfrac{\pi}{2})-c(\theta)\right)\,. \tag{27}$$

For $c(\theta)=c+\delta c\cos\theta$ this yields

$$\nu(\theta+\tfrac{\pi}{2})-\nu(\theta)=-\frac{n}{2L}(\sin\theta+\cos\theta)\delta c\,, \tag{28}$$

which reproduces (12).

Interpretation within the Ether Theory

We can use the above calculations and just replace the speed of light by its value (4) given within ether theory, that is, we use

$$c_+=c'(\theta,v)\qquad\text{and}\qquad c_-=c'(\theta+\pi,v) \tag{29}$$

and obtain for the observed frequency

$$\nu(\theta',v)=\frac{nc}{2L}\frac{1-\frac{v^2}{c^2}}{\sqrt{1-\frac{v^2}{c^2}(1-\cos^2\vartheta')}}\approx\frac{nc}{L}\left(1-\frac{1}{2}\frac{v^2}{c^2}\left(1+\cos^2\vartheta'\right)\right)\,, \tag{30}$$

what corresponds to the time light needs to propagate back and forth an interferometer arm. Comparison with the model independent calculation again gives a relation between the velocity with respect to the ether and the orientation dependence of the velocity of light

$$\frac{1}{2}\frac{v^2}{c^2}=\frac{\delta c}{c}\,. \tag{31}$$

For two orthogonally oriented resonators we obtain from (30) for the beat frequency

$$\begin{aligned}\nu(\theta'+\tfrac{\pi}{2},v)-\nu(\theta',v)&=\frac{nc}{2L}\frac{1-\frac{v^2}{c^2}}{\sqrt{1-\frac{v^2}{c^2}(1-\sin^2\theta')}}-\frac{nc}{2L}\frac{1-\frac{v^2}{c^2}}{\sqrt{1-\frac{v^2}{c^2}(1-\cos^2\vartheta')}}\\&=\frac{nc}{4L}\frac{v^2}{c^2}\cos(2\vartheta')+\mathcal{O}(v^4/c^4)\,,\end{aligned} \tag{32}$$

which is sensitive to the same quantity as interference experiments. In this case the comparison with the model independent calculation gives

$$\frac{1}{2}\frac{v^2}{c^2}=\sqrt{2}\frac{\delta c}{c}\,. \tag{33}$$

3.2 Constancy of Speed of Light

This class of experiments explore whether the outcome of experiments depends, via a velocity-dependent speed of light, on the velocity of the laboratory. As for the isotropy, this has been tested with interferometers as well as with resonators.

Interference Experiments

The Setup

The setup is essentially the same as for the Michelson–Morley experiment. The only difference is that we need unequal interferometer arm lengths, see Fig. 5. In the course of the experiment one varies the state of motion (velocity) of the apparatus and looks for associated variations in the intensity of the interfering light rays. For simplicity, we assume the interferometer arms do be orthogonal.

Model Independent Description

The intensity for an unequal arm Michelson interferometer has been given in (12). Now we assume that the speed of light may possibly depend on the velocity of the apparatus, too. The velocity is measured with respect to some inertial system. The result will not depend on the choice of this system.

We assume that δc may depend on the state of motion, too

$$\delta c_1 = \delta c(\theta, v), \quad \delta c_2 = \delta c(\theta + \tfrac{\pi}{2}, v)\,. \tag{34}$$

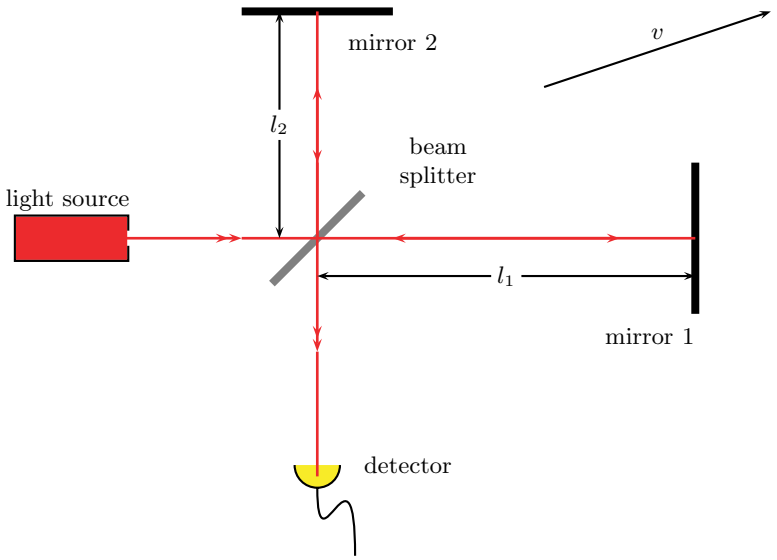


Fig. 5. The experiment of Kenndy and Thorndike uses a Michelson interferometer with different arm lengths $l_1 \neq l_2$

Then (12) yields

$$\Delta\phi(\theta, v) = \omega \left(\frac{l_1 - l_2}{c} + \frac{l_2}{c} \frac{\delta c(\theta + \frac{\pi}{2}, v)}{c} - \frac{l_1}{c} \frac{\delta c(\theta, v)}{c} \right) . \tag{35}$$

In the case that the speed of light does not depend on the direction this simplifies

$$\Delta\phi(v) = \omega \frac{\Delta l}{c} \left(1 + \frac{\delta c(v)}{c} \right) , \tag{36}$$

where $\Delta l = l_1 - l_2$. Therefore, a velocity dependence of the speed of light can be detected only if the interferometer arms are of unequal length.

What has been searched for in this type of experiments is the variation of the intensity while varying the velocity,

$$\delta\Delta\phi = \Delta\phi(v + \delta v) - \Delta\phi(v) = \omega \frac{\Delta l}{c} \frac{\delta c(v + \delta v) - \delta c(v)}{c} = \omega \frac{l_y - l_x}{c} \frac{\delta_v c}{c} . \tag{37}$$

Interpretation within the Ether Theory

The same calculation as for the Michelson–Morley experiment gives for an interferometer with unequal arm lengths $l_1 \neq l_2$ the phase shift

$$\Delta\phi = \frac{2\omega}{c} \frac{1}{1 - \frac{v^2}{c^2}} \left(l_1 \sqrt{1 - \frac{v^2}{c^2} \left(1 - \cos^2\vartheta\right)} - l_2 \sqrt{1 - \frac{v^2}{c^2} \left(1 - \sin^2\vartheta\right)} \right) \tag{38}$$

$$= \frac{2\omega}{c} \left(l_1 - l_2 + \frac{v^2}{c^2} \frac{1}{2} \left(l_1 - l_2 + l_1 \cos^2\vartheta - l_2 \sin^2\vartheta \right) \right) + \mathcal{O}(v^4/c^4) . \tag{39}$$

A change in the velocity with respect to the ether should result in a phase shift. (This is also the case for Michelson–Morley experiments, but there the velocity term is connected with the orientation which obscures a unique interpretation. Here the effect is related to a different arm length.) For a change of the velocity $\boldsymbol{v} \to \boldsymbol{v} + \delta\boldsymbol{v}$ we obtain from (39) to first order in the variation $\delta\boldsymbol{v}$

$$\delta\phi = \frac{2\omega}{c} \frac{\boldsymbol{v} \cdot \delta\boldsymbol{v}}{c^2} \left(l_1 - l_2 + l_1 \cos^2\vartheta - l_2 \sin^2\vartheta \right) . \tag{40}$$

For a given $\delta\boldsymbol{v}$ and a measured phase shift $\delta\phi$ one can conclude the value of $\boldsymbol{v}$. The larger the variation of the velocity, the better estimates will be.

A comparison with the model independent calculation gives

$$\frac{\delta_v c}{c} = 2 \frac{v}{c} \frac{\delta v}{c} . \tag{41}$$

Experiments with Resonators

The Setup

The setup is the same as described above. The only difference is that the setup will not be rotated but will change its state of motion. While changing the

velocity of he setup, one looks for a change of the frequency of the outcoupled electromagnetic wave. The variation of the state of motion of the laboratory is provided by the rotation of the Earth or its motion around the sun. For using the latter one has to use long term stable resonators.

Here we have to add an important remark of caution. The measurement of the frequency consists of a comparison of two frequencies, one frequency is given by the outcoupled wave, the other by some frequency standard. The frequency standard is defined by some atomic or molecular transition, for example. In principle, the frequency standard may also depend on its state of motion. This means that the present experiment explores whether two frequency standards, one given by the resonator, the other given by some atom or molecular transition, depend in the same or in a different way on the state of motion. In any case, as above any change of the measured frequency is, by convention, assigned to a change of the velocity of the light. Any definite statement regarding the 'true' cause of a (hypothetical) dependence of the signal from the state of motion can be made only by using a dynamical theory.

Model Independent Description

The frequency of the electromagnetic wave in the resonator is again given by (26) with the only modification that now the speed of light may depend on the state of motion of the apparatus, too, $c = c(\theta, v)$. Then

$$\nu(v,\theta) = \frac{n}{2L} c(\theta, v)\,. \tag{42}$$

A variation of the state of motion shows up in a variation of the measured frequency,

$$\delta\nu = \nu(v+\delta v,\theta) - \nu(v,\theta) = \frac{n}{2L}\left(c(v+\delta v,\theta) - c(v,\theta)\right) = \nu(v,\theta)\frac{\delta_v c(v,\theta)}{c}\,. \tag{43}$$

If no effect can be seen then $\delta_v c = 0$ within limits given by the accuracy of the apparatus. Also in this case one cannot distinguish a variation of the speed of light from a velocity dependent variation of the length of the resonator.

Interpretation within the Ether Theory

A change of the velocity results with (30) in the frequency shift

$$\begin{aligned}
\delta\nu &= \nu(\theta', v+\delta v) - \nu(\theta', v) \\
&= \frac{nc}{2L}\frac{1 - \frac{(v+\delta v)^2}{c^2}}{\sqrt{1 - \frac{(v+\delta v)^2}{c^2}(1-\cos^2\vartheta')}} - \frac{nc}{2L}\frac{1 - \frac{v^2}{c^2}}{\sqrt{1 - \frac{v^2}{c^2}(1-\cos^2\vartheta')}} \\
&\approx \frac{nc}{L}\,\frac{\boldsymbol{v}\cdot\delta\boldsymbol{v}}{c^2}\,\left(1+\cos^2\vartheta'\right)\,,
\end{aligned} \tag{44}$$

and a comparison with the model independent calculation gives

$$\frac{\delta_v\nu}{\nu} = \frac{\delta_v c}{c} = \frac{\boldsymbol{v}\cdot\delta\boldsymbol{v}}{c^2}\,. \tag{45}$$

4 The General Frame for Kinematical Test Theories

4.1 The Setting

In our kinematical test theory the consequences of transformations

$$t' = t'(t, \boldsymbol{x})\,, \qquad \boldsymbol{x}' = \boldsymbol{x}'(t, \boldsymbol{x}) \tag{46}$$

between the time and spatial coordinates of two observers are analyzed. On physical grounds we restrict to transformations which obey the following three requirements: The transformation

1. maps a force-free motion into a force-free motion, that is,

$$\frac{d^2\boldsymbol{x}}{dt^2} = 0 \quad \Leftrightarrow \quad \frac{d^2\boldsymbol{x}'}{dt'^2} = 0\,, \tag{47}$$

2. is a one-to-one mapping, and
3. the mapping depends on the relative velocity between the two observers only.

The first requirement implies a projective transformation [19] which with the second requirement gives the linearity of the transformation. From the third requirement we conclude that the linear transformation must have the particular structure

$$t' = a(v)t + e(v)\boldsymbol{v}\cdot\boldsymbol{x} \tag{48}$$

$$\boldsymbol{x}' = d(v)\boldsymbol{x} + b(v)\frac{\boldsymbol{v}(\boldsymbol{v}\cdot\boldsymbol{x})}{v^2} + f(v)\boldsymbol{v}t\,, \tag{49}$$

with undetermined function $a(v)$, $b(v)$, $d(v)$, $e(v)$, and $f(v)$. One function can be fixed by specifying the relative velocity between the observers and one function is related to the synchronization. Only three functions are of true physical nature and are related to the outcome of experiments.

The essential assumption now is that there exist a preferred frame Σ with coordinates $\boldsymbol{X}$ and T. In this frame light it assumed to propagate isotropically

$$ds^2 = dT^2 - dX^2 - dY^2 - dZ^2 = 0\,, \tag{50}$$

Usually, this preferred frame is identified with the cosmological frame in which the microwave background radiation is isotropic.

4.2 The General Transformation

The transformation between the preferred frame and another frame S with coordinates $(t', \boldsymbol{x})$ is described through (48,49)

$$t' = a(v)T + e(v)\boldsymbol{v}\cdot\boldsymbol{X} \tag{51}$$

$$\boldsymbol{x} = d(v)\boldsymbol{X} + b(v)\frac{\boldsymbol{v}(\boldsymbol{v}\cdot\boldsymbol{X})}{v^2} + f(v)\boldsymbol{v}T\,. \tag{52}$$

The velocity $\boldsymbol{v}$ between S and Σ is defined by the trajectory of the origin of S with respect to Σ, that is, $\boldsymbol{x} = 0$ is given by $\boldsymbol{X} = \boldsymbol{v}T$. That means

$$f(v) = -b(v) - d(v)\,. \tag{53}$$

Then we obtain the transformation

$$T = \frac{1}{a(v)}\left(t' - e(v)\boldsymbol{v}\cdot\boldsymbol{x}\right) \tag{54}$$

$$\boldsymbol{X} = \frac{1}{d(v)}\boldsymbol{x} - \left(\frac{1}{d(v)} - \frac{1}{b(v)}\right)\frac{\boldsymbol{v}(\boldsymbol{v}\cdot\boldsymbol{x})}{v^2} - \frac{\boldsymbol{v}}{a(v)}\left(t' - \boldsymbol{v}e(v)(\boldsymbol{v}\cdot\boldsymbol{x})\right)\,. \tag{55}$$

We have the freedom to introduce in S' another synchronization through $t = t' + \boldsymbol{\epsilon}'\cdot\boldsymbol{x}$. The coordinates in the corresponding system S are denoted by $(t, \boldsymbol{x})$. As a result, we obtain the transformations between Σ and S with arbitrary synchronization

$$T = \frac{1}{a(v)}\left(t - \boldsymbol{\epsilon}\cdot\boldsymbol{x}\right) \tag{56}$$

$$\boldsymbol{X} = \frac{1}{d(v)}\boldsymbol{x} - \left(\frac{1}{d(v)} - \frac{1}{b(v)}\right)\frac{\boldsymbol{v}(\boldsymbol{v}\cdot\boldsymbol{x})}{v^2} - \frac{1}{a(v)}\boldsymbol{v}(\boldsymbol{\epsilon}\cdot\boldsymbol{x}) + \frac{1}{a(v)}\boldsymbol{v}t\,, \tag{57}$$

where

$$\boldsymbol{\epsilon} := e(v)\frac{\boldsymbol{v}}{v} + \boldsymbol{\epsilon}'\,. \tag{58}$$

The line element in S comes out as

$$T^2 - \boldsymbol{X}^2 = \frac{1-v^2}{a^2}t^2 - 2\left(\frac{1-v^2}{a^2}\boldsymbol{\epsilon} + \frac{1}{ab}\boldsymbol{v}\right)\cdot\boldsymbol{x}\,t \tag{59}$$
$$-\frac{x^2}{d^2} + \frac{1-v^2}{a^2}(\boldsymbol{\epsilon}\cdot\boldsymbol{x})^2 + \frac{2}{ab}(\boldsymbol{v}\cdot\boldsymbol{x})(\boldsymbol{\epsilon}\cdot\boldsymbol{x}) + \left(\frac{1}{d^2} - \frac{1}{b^2}\right)\frac{(\boldsymbol{v}\cdot\boldsymbol{x})^2}{v^2}\,.$$

The light cone in S is defined by the vanishing of (59). We denote by θ the angle between the direction of light propagation and $\boldsymbol{v}$ and by θ' the angle between the speed of light and $\boldsymbol{\epsilon}$. Then in S the modulus of the speed of light

$$c(\theta, v, \epsilon) = \frac{|\boldsymbol{x}|}{t} = \tag{60}$$
$$\frac{bd(1-v^2)}{adv\cos\theta + bd\epsilon(1-v^2)\cos\theta' - a\sqrt{b^2(1-v^2) + (d^2 - b^2(1-v^2))\cos\theta}}$$

depends on the direction, on the velocity of the observer system and on the synchronization. This velocity will be used to describe the Michelson–Morley, Kennedy–Thorndike, and Ives-Stilwell experiments.

For later use we note

$$\lim_{v\to 0} a(v) = 1\,, \quad \lim_{v\to 0} b(v) = 1\,, \quad \lim_{v\to 0} d(v) = 1\,, \quad \lim_{v\to 0} \boldsymbol{\epsilon}(v) = 0\,, \tag{61}$$

what can be inferred from the property $t \to T$ and $\boldsymbol{x} \to \boldsymbol{X}$ for $\boldsymbol{v} \to 0$ in (56) and (56).

Special Relativity with arbitrary synchronization is characterized by

$$a(v) = \sqrt{1 - v^2}\,, \quad d(v) = 1\,, \quad b(v) = \frac{1}{\sqrt{1 - v^2}}\,, \tag{62}$$

and for standard Einstein synchronization we have in addition

$$\boldsymbol{\epsilon} = \boldsymbol{v}\,. \tag{63}$$

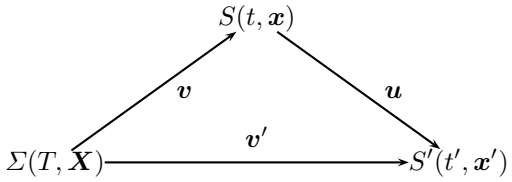
Then (56,57) give the Lorentz–transformations

$$T = \frac{1}{\sqrt{1 - v^2}}\,(t + \boldsymbol{v} \cdot \boldsymbol{x}) \tag{64}$$

$$\boldsymbol{X} = \boldsymbol{x}_\perp + \frac{1}{\sqrt{1 - v^2}}\,\left(\boldsymbol{x}_\parallel + \boldsymbol{v}t\right)\,. \tag{65}$$

4.3 Addition of Velocities

For the description of clock transport and the time dilation effects we need the addition of velocities in our general frame. We have three systems Σ, S and S' with corresponding relative velocities



(66)

The task is to represent $\boldsymbol{v}'$ as function of $\boldsymbol{v}$ and $\boldsymbol{u}$.

For that we insert $\boldsymbol{x}' = 0$ into the transformation $\Sigma \to S'$, and $\boldsymbol{x} = \boldsymbol{u}t$ into the transformation $\Sigma \to S$. Elimination of T and $\boldsymbol{X}$ gives

$$\frac{1}{a(v')}t' = \frac{1}{a(v)}\,(t - \boldsymbol{\epsilon} \cdot \boldsymbol{u}t) \tag{67}$$

$$\frac{1}{a(v')}\boldsymbol{v}'t' = \frac{1}{d(v)}\boldsymbol{u}t - \left(\frac{1}{d(v)} - \frac{1}{b(v)}\right)\frac{\boldsymbol{v}(\boldsymbol{v} \cdot \boldsymbol{u}t)}{v^2} - \frac{1}{a(v)}\boldsymbol{v}(\boldsymbol{\epsilon} \cdot \boldsymbol{u}t) + \frac{1}{a(v)}\boldsymbol{v}t \tag{68}$$

which yields

$$\boldsymbol{v}' = \boldsymbol{v} + \frac{\dfrac{a(v)}{d(v)}\boldsymbol{u}_\perp + \dfrac{a(v)}{b(v)}\boldsymbol{u}_\parallel}{1 - \boldsymbol{\epsilon} \cdot \boldsymbol{u}}\,. \tag{69}$$

For the choice (62) of the parameters we obtain the special relativistic expression. For small velocities $\boldsymbol{u}$,

$$\boldsymbol{v}' \approx \boldsymbol{v} + \frac{a(v)}{d(v)}\boldsymbol{u} - \left(\frac{a(v)}{d(v)} - \frac{a(v)}{b(v)}\right)\frac{\boldsymbol{v}(\boldsymbol{v} \cdot \boldsymbol{u})}{v^2}\,. \tag{70}$$

5 The Test Theory of Robertson

The test theory of Robertson [8] now specializes the above formalism to the special case of Einstein synchronization. The resulting theory is physically equivalent to the original one.

5.1 The Einstein–synchronization

In order to determine the coefficient $\boldsymbol{\epsilon}$ for the Einstein–synchronization we consider two clocks A and B which are at rest in a system S. This system S moves with a velocity $\boldsymbol{v}$ with respect to Σ. At $t = 0$, a signal is sent from A an arrives in B at $t = t_1$. This signal will be sent back immediately and reaches A at t_2, see Fig. 6. Einstein synchronization now requires (compare, e.g., [9])

$$t_2 = 2\,t_1\,. \tag{71}$$

According to the diagram (66) and the relations (56,57) we represent the events E_1 and E_2 in the relations between S and Σ as well as in the relations between S' and Σ. Since the clock A is at rest in the moving system S, we have $\boldsymbol{x}_2 = 0$ and $\boldsymbol{X}_2 = \boldsymbol{v}T_2$. Therefore,

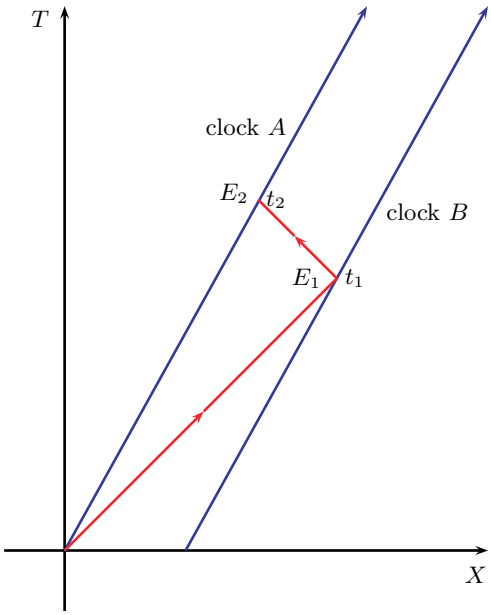


Fig. 6. The Einstein–synchronization: A and B are worldlines of two clocks at rest in a system which moves with respect to Σ. At $t = 0$, the observer A sends a light signal to B, where it arrives at time t_1. The signal sent back immediately arrives in A at time t_2. The Einstein–synchronization now requires $t_1 = \frac{1}{2}t_2$

$$T_2 = \frac{1}{a(v)} t_2 \qquad \boldsymbol{X}_2 = \boldsymbol{v} T_2 \,. \tag{72}$$

From the equations for light propagation

$$|\boldsymbol{X}_1|^2 = T_1^2 \,, \qquad |\boldsymbol{X}_2 - \boldsymbol{X}_1|^2 = (T_2 - T_1)^2 \,, \tag{73}$$

we then obtain after some calculations

$$\boldsymbol{\epsilon} = -\boldsymbol{v} \frac{a(v)}{b(v)(1-v^2)} \,. \tag{74}$$

Using this condition we obtain for our transformations (56,57)

$$T = \frac{1}{a(v)} \left(t + \frac{a(v)}{b(v)(1-v^2)} \boldsymbol{v} \cdot \boldsymbol{x} \right) \tag{75}$$

$$\boldsymbol{X} = \frac{1}{d(v)} \boldsymbol{x}_\perp + \frac{1}{b(v)(1-v^2)} \boldsymbol{x}_\parallel^2 + \frac{1}{a(v)} \boldsymbol{v} t \,, \tag{76}$$

where $\boldsymbol{x}_\perp = \boldsymbol{x} - \boldsymbol{x}_\parallel$ with $\boldsymbol{x}_\parallel = \dfrac{\boldsymbol{v} \cdot \boldsymbol{x}}{v^2} \boldsymbol{v}$. The line element S will be

$$T^2 - \boldsymbol{X}^2 = \frac{1-v^2}{a^2(v)} t^2 - \frac{1}{d^2(v)} \boldsymbol{x}_\perp^2 - \frac{1}{b^2(v)\,(1-v^2)} \boldsymbol{x}_\parallel^2 \tag{77}$$

which has the structure

$$ds^2 = g_0^2(v) t^2 - \left(g_1^2(v) d\boldsymbol{x}_\parallel^2 + g_2^2(v) d\boldsymbol{x}_\perp^2 \right) \tag{78}$$

with

$$g_0^2(v) = \frac{1-v^2}{a^2(v)}, \quad g_1^2(v) = \frac{1}{b^2(v)(1-v^2)}, \quad g_2^2(v) = \frac{1}{d^2(v)} \,. \tag{79}$$

This is (59) in the case of the Einstein–synchronization. The measurements of length depend on the velocity of the frame which violates the relativity principle. For SR $g_0(v) = g_1(v) = g_2(v) = 1$ for all $\boldsymbol{v}$.

The speed of light is

$$c(\theta, v) = \frac{1}{B(v)} \frac{1}{\sqrt{1 + A^2(v) \cos^2 \theta}} \,, \tag{80}$$

where we defined the modulus $1/B(v)$ and the anisotropy $A(v)$

$$\frac{1}{B(v)} = \frac{d(v)\sqrt{1-v^2}}{a(v)} = \frac{g_0(v)}{g_2(v)} \tag{81}$$

$$A(v) = \frac{d(v)}{b(v)\sqrt{1-v^2}} - 1 = \frac{\sqrt{g_1^2(v) - g_2^2(v)}}{g_2(v)} \,. \tag{82}$$

In the subspace orthogonal to $\boldsymbol{v}$ the speed of light is isotropic. The relative velocity $\boldsymbol{v}$ defines a preferred direction. The function $a(v)$ is not related to a possible anisotropy of c. If we define the speed of light in direction of and orthogonal to $\boldsymbol{v}$

$$c_{\parallel}(v) = c(0,v) = \frac{b(v)(1-v^2)}{a(v)} = \frac{\sqrt{1-v^2}}{A(v)} \tag{83}$$

$$c_{\perp}(v) = c(\tfrac{\pi}{2},v) = \frac{1}{a(v)} d(v)\sqrt{1-v^2} = \frac{1}{B(v)} \tag{84}$$

then

$$c(\theta,v) = \frac{c_{\perp}(v)}{\sqrt{1+\frac{c_{\perp}^2(v)-c_{\parallel}^2(v)}{c_{\parallel}^2(v)}\cos^2\theta}}\,. \tag{85}$$

Therefore the anisotropy is given by the relative difference of $c_{\parallel}$ and $c_{\perp}$.

Using $c_{\parallel}$ and $c_{\perp}$ we also can express the transformations (56,57)

$$T = \frac{1}{a(v)}\,(t - \boldsymbol{\epsilon}\cdot\boldsymbol{x}) \tag{86}$$

$$\begin{aligned}\boldsymbol{X} &= \frac{\sqrt{1-v^2}}{a(v)}\left(\frac{\boldsymbol{x}_{\perp}}{c_{\perp}} + \frac{\sqrt{1-v^2}}{c_{\parallel}}\boldsymbol{x}_{\parallel} - \frac{\boldsymbol{v}\,(t-\boldsymbol{\epsilon}\cdot\boldsymbol{x})}{\sqrt{1-v^2}}\right)\\ &= \frac{\sqrt{1-v^2}}{a(v)}\left(B(v)\boldsymbol{x}_{\perp} + A(v)\boldsymbol{x}_{\parallel} - \frac{\boldsymbol{v}\,(t-\boldsymbol{\epsilon}\cdot\boldsymbol{x})}{\sqrt{1-v^2}}\right),\end{aligned} \tag{87}$$

and the line element (59)

$$\begin{aligned}T^2 - \boldsymbol{X}^2 = \frac{1-v^2}{a^2(v)}\Bigg(t^2 - 2\left(\boldsymbol{\epsilon} + \frac{\boldsymbol{v}}{c_{\parallel}}\right)\cdot\boldsymbol{x}\,t - \frac{1}{c_{\perp}^2}x^2\\ + \left(\left(\boldsymbol{\epsilon} + \frac{\boldsymbol{v}}{c_{\parallel}}\right)\cdot\boldsymbol{x}\right)^2 + \left(\frac{1}{c_{\perp}^2} - \frac{1}{c_{\parallel}^2}\right)\frac{(\boldsymbol{v}\cdot\boldsymbol{x})^2}{v^2}\Bigg)\,.\end{aligned} \tag{88}$$

Here we like to add some remarks on other synchronizations. We show (i) that the synchronization by slow-clock transport yields a different $\boldsymbol{\epsilon}$ and (ii) that the requirement of coincidence of Einstein with slow-clock transport synchronization is only possible for $a(v) = \sqrt{1-v^2}$ [9].

For slow–clock synchronization we consider a clock moving with a small velocity with respect to the system S. By passing the clocks at rest in S, these clocks will be given the time of the slowly moving clock, see Fig. 7. Since the moving clocks are at rest in S' we have $T = t'/a(v')$. The same clocks is described in S by $T = (t - \boldsymbol{\epsilon}\cdot\boldsymbol{x})t$. The prescription of synchronization by slow clock transport now is $t' = t$ from which we immediately obtain $\boldsymbol{\epsilon}\cdot\boldsymbol{x} = (a(v) - a(v'))T$.

Furthermore, from $\boldsymbol{X} = \boldsymbol{v}'T$ in (57) and the addition of velocities (70) for small $\boldsymbol{u}$ we obtain $a(v)(\boldsymbol{v}\cdot\boldsymbol{u})T = \boldsymbol{v}\cdot\boldsymbol{x}$. Then

$$\begin{aligned}\boldsymbol{\epsilon}\cdot\boldsymbol{x} &= -\frac{1}{a(v)}\left(a(v') - a(v)\right)\frac{\boldsymbol{v}\cdot\boldsymbol{x}}{\boldsymbol{u}\cdot\boldsymbol{v}}\\ &\approx -\frac{1}{a(v)}\left((\boldsymbol{v}'-\boldsymbol{v})\cdot\boldsymbol{\nabla}_v a(v)\right)\frac{\boldsymbol{v}\cdot\boldsymbol{x}}{\boldsymbol{u}\cdot\boldsymbol{v}}\\ &= -\frac{1}{b(v)}\frac{1}{v}\frac{da(v)}{dv}\boldsymbol{v}\cdot\boldsymbol{x}\,,\end{aligned} \tag{89}$$

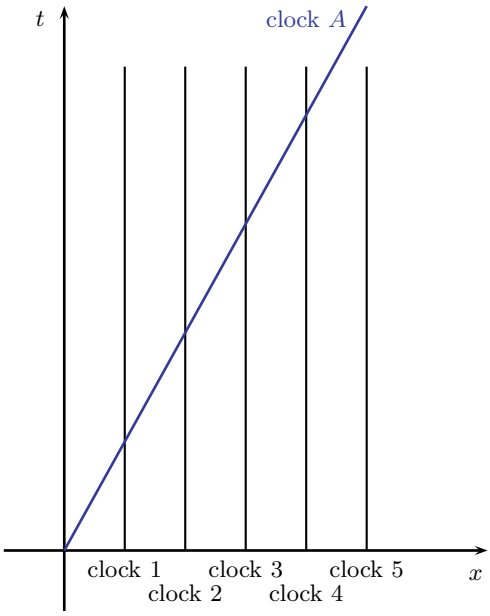


Fig. 7. Synchronization by using slow clocks: a clock A moves slowly in S and sets all clocks in S at its own time

where we again used (70) for small $\boldsymbol{u}$. Since this should hold for all $\boldsymbol{x}$ we obtain for the synchronization parameter

$$\boldsymbol{\epsilon} = \frac{1}{b(v)} \frac{da(v)}{dv} \frac{\boldsymbol{v}}{v} \,. \tag{90}$$

This is different from the result (74) for the Einstein synchronization.

If we require that both methods of synchronization should lead to the same synchronization parameter

$$\frac{da(v)}{dv} = -\frac{v}{(1-v^2)} a(v) \,, \tag{91}$$

we obtain, after integration,

$$a(v) = \sqrt{1-v^2} \,. \tag{92}$$

This is the ordinary time dilation factor. Only in this case both synchronization schemes coincide [9].

5.2 Discussion of the Experiments

Based on (78) we discuss the three classes of experiments, namely the experiments testing the isotropy of the speed of light (Michelson–Morley-experiments), the experiments testing the independence of the speed of light from the velocity of the apparatus (Kennedy–Thorndike-experiments) and the experiments measuring the time dilatation in terms of the Doppler effect (Ives–Stilwell–experiments). These three experiments together imply the Lorentz–transformations and, thus, Lorentz invariance.

Isotropy of the Speed of Light

For a Michelson–Morley experiment with interferometer arms of equal length l the general phase shift (11) with (80) yields

$$\delta\phi = 2\omega l \left(\frac{1}{c_2} - \frac{1}{c_1}\right) = 2\frac{\omega l}{B(v)} \left(\sqrt{1 + A(v)\sin^2\theta} - \sqrt{1 + A(v)\cos^2\theta}\right) . \quad (93)$$

This is independent from the orientation only if the anisotropy $A(v)$ vanishes,

$$A(v) = 0 \quad \Leftrightarrow \quad g_1(v) = g_2(v) \quad \Leftrightarrow \quad d(v) = b(v)\sqrt{1 - v^2} \quad (94)$$

The effect does not depend on the time dilation factor g_0. As a consequence the speed of light (80) reduces to

$$c(\theta, v) = c(v) = 1/B(v) , \quad (95)$$

what still may depend on v. In principle, these experiments have to carried through for all v.

For the description of experiments with resonators we use (26) and (80) and obtain

$$\nu(\theta, v) = \frac{n}{2L} c(\theta, v) = \frac{n}{2LB(v)} \frac{1}{\sqrt{1 + A(v)\cos^2\theta}} . \quad (96)$$

This again does not depend on the orientation if (94) holds.

In the case of a setup with two orthogonally oriented resonators the relative change of the two frequencies is

$$\frac{\nu(\vartheta + \frac{\pi}{2}, v) - \nu(\vartheta, v)}{\nu(\frac{\pi}{2}, v)} = \sqrt{\frac{1 + A(v)\cos^2\vartheta}{1 + A(v)\sin^2\vartheta}} - 1 . \quad (97)$$

The lack of any signal again yields (94).

Constancy of the Speed of Light

Here we take the isotropy of the speed of light as granted, that is, we assume (94).

For an unequal arm interferometer we obtain from (11) and (95)

$$\delta\phi = 2\omega \left(\frac{l_2}{c_2} - \frac{l_1}{c_1}\right) = 2\frac{\omega(l_2 - l_1)}{B(v)} . \quad (98)$$

This phase shift is independent from the velocity v of the apparatus if $B(v) = K$, that is, $g_0(v) = Kg_1(v)$, where K is some constant. The condition $\lim_{v\to 0} g_1(v) = \lim_{v\to 0} g_0(v) = 1$ implies $K = 1$. As a consequence

$$B(v) = 1 \quad \Leftrightarrow \quad g_0(v) = g_1(v) . \quad (99)$$

Fig. 8. Example for isotropic (*left*) and anisotropic (*right*) propagation

Also the frequency of the radiation outcoupled from a resonator

$$\nu(v) = \frac{n}{2L}c(v) = \frac{n}{2LB(v)} \tag{100}$$

does not depend on the velocity of the resonator if (99) holds.

The independence from the orientation and the velocity of the apparatus yields with (79)

$$b^2(v) = \frac{a^2(v)}{(1-v^2)^2} \qquad \text{and} \qquad d(v) = \frac{a(v)}{\sqrt{1-v^2}}\,. \tag{101}$$

or, equivalently,

$$g_0(v) = g_1(v) = g_2(v) \tag{102}$$

In this case the line element is $ds^2 = t^2 - \boldsymbol{x}^2 = 0$. The function $a(v)$ is the only unknown function remaining in the transformations (75,76)

$$T = \frac{1}{a(v)}\left(t + \boldsymbol{v}\cdot\boldsymbol{x}\right) \tag{103}$$

$$\boldsymbol{X} = \frac{\sqrt{1-v^2}}{a(v)}\left(\boldsymbol{x}_\perp + \frac{1}{\sqrt{1-v^2}}\left(\boldsymbol{x}_\parallel + \boldsymbol{v}t\right)\right)\,. \tag{104}$$

Time Dilation

We still need a further experiment which can determine the remaining function $g_0(v)$ or $a(v)$ which gives the time dilatation. Such an experiment is the Doppler shift, for example. In these experiment the frequency of radiation emitted from moving sources will be measured. In this setup, both the laboratory as well as the source will move with respect to the preferred frame. Therefore we need the transformations between frames S and S' moving with v and v' with respect to Σ. This transformation can be easily derived and reads

$$t' = \frac{a(v')(1+\boldsymbol{u}\cdot\boldsymbol{v})}{a(v)(1-u^2)}\left(t - \boldsymbol{x}\cdot\boldsymbol{u}\right) \tag{105}$$

$$\boldsymbol{x}' = \frac{a(v')(1+\boldsymbol{u}\cdot\boldsymbol{v})}{a(v)\sqrt{1-u^2}}\left(\boldsymbol{x} - \left(1 - \frac{1}{\sqrt{1-u^2}}\right)\frac{\boldsymbol{u}(\boldsymbol{u}\cdot\boldsymbol{x})}{u^2} + \frac{\boldsymbol{u}}{\sqrt{1-u^2}}t\right), \tag{106}$$

where $\boldsymbol{u}$ is the velocity of S' with respect to S.

Since experiments on time dilation measure the relation between t and t' one can determine the last unknown function $g_0(v)$ or $a(v)$. For doing so we observe the frequency of a radiating atom moving with velocity $\boldsymbol{u}$ in the laboratory S or, equivalently, with velocity $\boldsymbol{v}'$ with respect to Σ. We define a system S' in which the atom is at rest. For the determination of $a(v)$ it is enough to have $\boldsymbol{v}$, $\boldsymbol{v}'$, and thus $\boldsymbol{u}$, in x-direction. S and S' are related to Σ via (103,104) $\boldsymbol{v}$ and $\boldsymbol{v}'$ as relative velocities.

We emphasize that we do know neither the magnitude not the direction of the velocity $\boldsymbol{v}$ with respect to the preferred system Σ. Therefore we carry through the following calculations in full generality.

The Doppler Formula

In the preferred frame Σ light rays obey the usual relation

$$(T_\mathrm{r} - T_\mathrm{s})^2 = |\boldsymbol{X}_\mathrm{r} - \boldsymbol{X}_\mathrm{s}|^2 \tag{107}$$

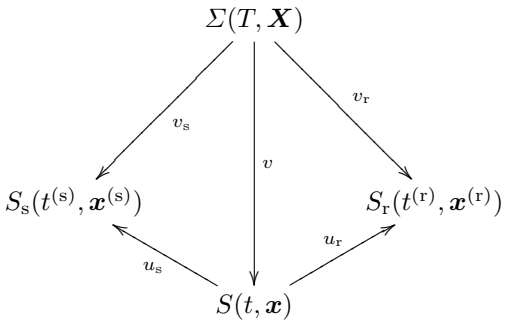
With (103,104) we can transform this to relations for the coordinates in S and obtain

$$t_\mathrm{rs} = x_\mathrm{rs}\,, \tag{108}$$

where we defined

$$t_\mathrm{rs} = t_\mathrm{r} - t_\mathrm{s}\,, \quad x_\mathrm{rs} = |\boldsymbol{x}_\mathrm{rs}|\,, \quad \boldsymbol{x}_\mathrm{rs} = \boldsymbol{x}_\mathrm{r} - \boldsymbol{x}_\mathrm{s} \tag{109}$$

Now we consider the situation shown in the diagram



(110)

and two light rays emitted at events $(t_\mathrm{s}^{(1)}, \boldsymbol{x}_\mathrm{s}^{(1)})$ and $(t_\mathrm{s}^{(2)}, \boldsymbol{x}_\mathrm{s}^{(2)})$ and received at $(t_\mathrm{r}^{(1)}, \boldsymbol{x}_\mathrm{r}^{(1)})$ and $(t_\mathrm{r}^{(2)}, \boldsymbol{x}_\mathrm{r}^{(2)})$, see Fig. 9. $\boldsymbol{u}_\mathrm{s}$ and $\boldsymbol{u}_\mathrm{r}$ are the velocities of the sender and receiver with respect to S. Then in S

$$\boldsymbol{x}_\mathrm{rs}^{(2)} - \boldsymbol{x}_\mathrm{rs}^{(1)} = \boldsymbol{x}_\mathrm{r}^{(2)} - \boldsymbol{x}_\mathrm{r}^{(1)} - \boldsymbol{x}_\mathrm{s}^{(2)} + \boldsymbol{x}_\mathrm{s}^{(1)} = \boldsymbol{u}_\mathrm{r}\Delta t_\mathrm{r} - \boldsymbol{u}_\mathrm{s}\Delta t_\mathrm{s}\,, \tag{111}$$

where $\Delta t_\mathrm{s} = t_\mathrm{s}^{(2)} - t_\mathrm{s}^{(1)}$ and $\Delta t_\mathrm{r} = t_\mathrm{r}^{(2)} - t_\mathrm{r}^{(1)}$. From (108) and (111) we obtain

$$\begin{aligned}\Delta t_\mathrm{r} &= t_\mathrm{s}^{(2)} + x_\mathrm{rs}^{(2)} - (t_\mathrm{s}^{(1)} + x_\mathrm{rs}^{(1)}) \\ &= \Delta t_\mathrm{s} + \boldsymbol{n}\cdot(\boldsymbol{x}_\mathrm{rs}^{(2)} - \boldsymbol{x}_\mathrm{rs}^{(1)}) \\ &= \Delta t_\mathrm{s} + \boldsymbol{n}\cdot(\boldsymbol{u}_\mathrm{r}\Delta t_\mathrm{r} - \boldsymbol{u}_\mathrm{s}\Delta t_\mathrm{s})\,,\end{aligned} \tag{112}$$

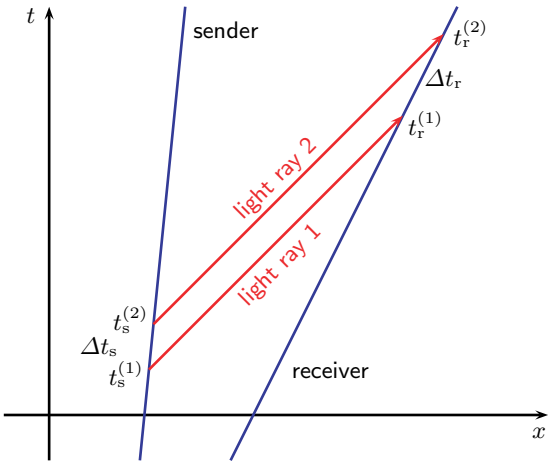


Fig. 9. The observed frequency $1/\Delta t_{\rm r}$ is function of the emitted frequency $1/\Delta t_{\rm s}$

where we used $|\boldsymbol{x}_{\rm rs}^{(2)}| = |\boldsymbol{x}_{\rm rs}^{(1)}| + (\boldsymbol{x}_{\rm rs}^{(2)} - \boldsymbol{x}_{\rm rs}^{(1)})\boldsymbol{\nabla}|\boldsymbol{x}_{\rm rs}^{(1)}|$ and $\boldsymbol{n} = \boldsymbol{x}_{\rm rs}/|\boldsymbol{x}_{\rm rs}|$. This gives

$$\Delta t_{\rm r} = \frac{1 - \boldsymbol{n}\cdot\boldsymbol{u}_{\rm s}}{1 - \boldsymbol{n}\cdot\boldsymbol{u}_{\rm r}}\Delta t_{\rm s}\,. \tag{113}$$

We still have to replace the coordinate times $\Delta t_{\rm r}$ and $\Delta t_{\rm s}$ by the corresponding eigentimes $\Delta t_{\rm r}^{(\rm r)}$ and $\Delta t_{\rm s}^{(\rm s)}$. For doing so we use (103)

$$T_{\rm s,r} = \frac{1}{a(v_{\rm s,r})}\left(t^{(\rm s,r)} + \boldsymbol{v}_{\rm s,r}\cdot\boldsymbol{x}^{(\rm s,r)}\right)\,, \tag{114}$$

where $\boldsymbol{v}_{\rm s}$ is the relative velocity of the sender with respect to the preferred frame Σ. Since the clock of the sender/receiver is at rest in $S_{\rm s,r}$ we have $\boldsymbol{x}_{\rm s,r} = 0$ and

$$\Delta T_{\rm s,r} = \frac{1}{a(v_{\rm s,r})}\Delta t_{\rm s,r}^{(\rm s,r)}\,. \tag{115}$$

We furthermore get from (103)

$$\Delta T_{\rm s,r} = \frac{1}{a(v)}\left(\Delta t_{\rm s,r} + \boldsymbol{v}\cdot\Delta\boldsymbol{x}_{\rm s,r}\right) = \frac{1}{a(v)}\left(1 + \boldsymbol{v}\cdot\boldsymbol{u}_{\rm s,r}\right)\Delta t_{\rm s,r}\,. \tag{116}$$

With that we can eliminate $\Delta T_{\rm s,r}$ and, thus, can express the eigentime $\Delta t_{\rm s,r}^{(\rm s,r)}$ by $\Delta t_{\rm s,r}$

$$\Delta t_{\rm s,r}^{(\rm s,r)} = \frac{a(v'_{\rm s,r})}{a(v)}\left(1 + \boldsymbol{v}\cdot\boldsymbol{u}_{\rm s,r}\right)\Delta t_{\rm s,r}\,. \tag{117}$$

In terms of the frequencies defined by $\nu_{\rm s,r} = 1/\Delta t_{\rm s,r}^{(\rm s,r)}$ we thus have

$$\frac{\nu_{\rm r}}{\nu_{\rm s}} = \frac{\Delta t_{\rm s}^{(\rm s)}}{\Delta t_{\rm r}^{(\rm r)}} = \frac{a(v_{\rm s})\left(1 + \boldsymbol{v}\cdot\boldsymbol{u}_{\rm s}\right)\left(1 - \boldsymbol{n}\cdot\boldsymbol{u}_{\rm r}\right)}{a(v_{\rm r})\left(1 + \boldsymbol{v}\cdot\boldsymbol{u}_{\rm r}\right)\left(1 - \boldsymbol{n}\cdot\boldsymbol{u}_{\rm s}\right)}\,. \tag{118}$$

We specialize to a receiver at rest in S, $\boldsymbol{u}_{\mathrm{r}} = 0$, and finally obtain the Doppler formula we need

$$\frac{\nu_{\mathrm{r}}}{\nu_{\mathrm{s}}} = \frac{a(v'_{\mathrm{s}})\,(1 + \boldsymbol{v}\cdot\boldsymbol{u}_{\mathrm{s}})}{a(v)(1 - \boldsymbol{n}\cdot\boldsymbol{u}_{\mathrm{s}})}\,. \tag{119}$$

The Experiment

If one measures the frequency emitted by a moving atom parallel and anti–parallel to the velocity of the atom, then the experiments gives (and this is also what SR predicts) that the product is just the square of the frequency of the atom at rest

$$\nu_{\mathrm{r}}^{+}\nu_{\mathrm{r}}^{-} = \nu_{\mathrm{s}}^{2}\,. \tag{120}$$

With (119) this means (for ν_{r}^{+} we chose the direction $\boldsymbol{n} = \boldsymbol{n}_0$, and for ν_{E}^{-} the direction $\boldsymbol{n} = -\boldsymbol{n}_0$)

$$\frac{\nu_{\mathrm{r}}^{+}}{\nu_{\mathrm{s}}}\frac{\nu_{\mathrm{r}}^{-}}{\nu_{\mathrm{s}}} = \frac{a(v'_{\mathrm{s}})\,(1 + \boldsymbol{v}\cdot\boldsymbol{u}_{\mathrm{s}})}{a(v)(1 - \boldsymbol{n}_0\cdot\boldsymbol{u}_{\mathrm{s}})}\frac{a(v'_{\mathrm{s}})\,(1 + \boldsymbol{v}\cdot\boldsymbol{u}_{\mathrm{s}})}{a(v)(1 + \boldsymbol{n}_0\cdot\boldsymbol{u}_{\mathrm{s}})} = \frac{a^2(v'_{\mathrm{s}})\,(1 + \boldsymbol{v}\cdot\boldsymbol{u}_{\mathrm{s}})^2}{a^2(v)(1 - u_{\mathrm{s}}^2)} = 1 \tag{121}$$

with $u_{\mathrm{s}} = \boldsymbol{n}_0\cdot\boldsymbol{u}_S$, so that

$$\frac{a(v'_{\mathrm{s}})(1 + \boldsymbol{v}\cdot\boldsymbol{u}_{\mathrm{s}})}{a(v)} = \sqrt{1 - u_{\mathrm{s}}^2}\,. \tag{122}$$

We abbreviate $\boldsymbol{v}_{\mathrm{s}} = \boldsymbol{v}$ and $\boldsymbol{u}_{\mathrm{s}} = \boldsymbol{u}$ and use this result in the transformations (105,106)

$$t' = \frac{1}{\sqrt{1 - u^2}}\,(t - \boldsymbol{x}\cdot\boldsymbol{u}) \tag{123}$$

$$\boldsymbol{x}' = \boldsymbol{x} - \left(1 - \frac{1}{\sqrt{1 - u^2}}\right)\frac{\boldsymbol{u}(\boldsymbol{u}\cdot\boldsymbol{x})}{u^2} + \boldsymbol{u}t\,. \tag{124}$$

This are the ordinary Lorentz–transformations. Any information about the state of motion with respect to Σ disappeared. The preferred frame Σ plays no role anymore.

We are also in the position to determine the function $a(v)$: All considerations above hold for all systems S and S', so that we can assume $\boldsymbol{v} = 0$ for a particular system S. Then, from (122) we infer $a(u) = \sqrt{1 - u^2}a(0)$. Since we should obtain the identity for $\boldsymbol{u} = 0$, compare (61), we finally obtain $a(0) = 1$. Therefore, with three experiments we were able to determine all three parameter functions $a(v)$, $b(v)$ and $d(v)$.

5.3 Linearization of the Robertson Test Theory

Since in all laboratory experiments the velocities are relatively small compared to the speed of light and since also the velocity of the Earth with respect to the Sun and the cosmological preferred frame is small, we can expand the functions $g_r(v)$ $(r = 1, 2, 3)$ with respect to the velocities

$$g_r(v) = 1 + \frac{1}{2} g_r^0 v^2 + \dots , \tag{125}$$

where we used (61). As a consequence, the determination of the functions $g_r(v)$ reduces to the determination of three parameters g_r^0.

In this approximation the important combinations are given by

$$\frac{g_2(v) - g_1(v)}{g_1(v)} = \frac{1}{2} \left(g_2^0 - g_1^0 \right) v^2, \quad \left| \frac{g_0(v)}{g_1(v)} - 1 \right| = \frac{1}{2} \left| g_0^0 - g_1^0 \right| v^2 \tag{126}$$

and $g_0(v) = 1 + \frac{1}{2} g_0^0 v^2$. In this approximation the Michelson–Morley experiment implies $g_1^0 = g_2^0$ and the Kennedy–Thorndike-experiment $g_0^0 = g_1^0$. The time dilation experiment yields $g_0^0 = 0$. The parameter combinations $g_2^0 - g_1^0$ and $g_0^0 - g_1^0$ will show up again in the Mansouri–Sexl test theory.

6 The General Formalism

Based on the transformations (56,57), the line element (59) and the corresponding speed of light (60) we describe now all experiments without the assumption of Einstein synchronization. We will show that again the three previous experiments are enough to characterize Lorentz invariance. However, the basic physical quantities will be slightly different, namely the two-way speed of light and the two-way Doppler shift.

6.1 The Frame

Though the one-way velocity of light depends on the synchronization parameter, the two-way velocity defined by

$$\frac{2}{c_{(2)}(v, \epsilon, \vartheta, \vartheta')} = \frac{1}{c(v, \epsilon, \vartheta, \vartheta')} + \frac{1}{c(v, \epsilon, \vartheta + \pi, \vartheta')} \tag{127}$$

has the same form as the one-way velocity of light under the assumption of Einstein–synchronization

$$c_{(2)}(v, \epsilon, \vartheta, \vartheta') = c_{(2)}(\vartheta, v) = \frac{1}{B(v)} \frac{1}{\sqrt{1 + A(v) \cos^2 \vartheta}} , \tag{128}$$

which no longer depends on $\boldsymbol{\epsilon}$.

Since the two-way velocity of light is exactly the same as the one-way velocity for Einstein synchronization, the results (101) for the experiments testing the isotropy and constancy of the speed of light are also the same and, thus, need not to be repeated. The only nontrivial experiment is the time dilation experiment.

We use the results (101) in order to eliminate $b(v)$ and $d(v)$ in (56,57)

$$T = \frac{1}{a(v)}\left(t - \boldsymbol{\epsilon}\cdot\boldsymbol{x}\right) \tag{129}$$

$$\boldsymbol{X} = \frac{\sqrt{1-v^2}}{a(v)}\left(\boldsymbol{x} - \left(1-\sqrt{1-v^2}\right)\frac{\boldsymbol{v}(\boldsymbol{v}\cdot\boldsymbol{x})}{v^2}\right) - \frac{1}{a(v)}\left(\boldsymbol{v}(\boldsymbol{\epsilon}\cdot\boldsymbol{x}) - \boldsymbol{v}t\right), \tag{130}$$

and in the expression for the light cone in S

$$0 = T^2 - \boldsymbol{X}^2 = \frac{1-v^2}{a^2(v)}\left(t^2 - 2\left(\boldsymbol{\epsilon}+\boldsymbol{v}\right)\cdot\boldsymbol{x}t - \boldsymbol{x}^2 + \left(\left(\boldsymbol{\epsilon}+\boldsymbol{v}\right)\cdot\boldsymbol{x}\right)^2\right), \tag{131}$$

which is the SR lightcone for arbitrary synchronization. Again, we have to determine the remaining function $a(v)$.

In order to discuss time dilation effects we again have to consider the transformation between two systems S and S' moving with $\boldsymbol{v}$ and $\boldsymbol{v}'$ with respect to the preferred frame Σ. Using (129,130), a lengthy calculation yields

$$t' = \frac{a(v')}{a(v)}\left(t - \boldsymbol{\epsilon}\cdot\boldsymbol{x}\right) + \boldsymbol{\epsilon}'\cdot\boldsymbol{x}' \tag{132}$$

$$\begin{aligned}\boldsymbol{x}' = \frac{a(v')\gamma'}{a(v)\gamma}\Bigg(&\boldsymbol{x} + \boldsymbol{v}\left(\gamma\left(t - \boldsymbol{\epsilon}\cdot\boldsymbol{x}\right) - \left(1-\frac{1}{\gamma}\right)\frac{(\boldsymbol{v}\cdot\boldsymbol{x})}{v^2}\right)\\ &+\boldsymbol{v}'\left(\frac{\gamma'-1}{v'^2}\left(\boldsymbol{v}'\cdot\boldsymbol{x} - \left(1-\frac{1}{\gamma}\right)\frac{(\boldsymbol{v}\cdot\boldsymbol{x})(\boldsymbol{v}'\cdot\boldsymbol{v})}{w^2} + (\boldsymbol{v}'\cdot\boldsymbol{v})\gamma\left(t - \boldsymbol{\epsilon}\cdot\boldsymbol{x}\right)\right)\right.\\ &\left.-\gamma'\gamma\left(t - \boldsymbol{\epsilon}\cdot\boldsymbol{x}\right)\right)\Bigg)\end{aligned} \tag{133}$$

where $\gamma' = \gamma(v')$.

6.2 Discussion of the Experiments: Time Dilatation

The Doppler Formula

Again we calculate and use the Doppler effect for moving atoms in order to determine the time dilatation. As before, we have to calculate the time differences between the emission and reception times of two light rays, see Fig. 9 and diagram (110). Again we can start from (107), use (129,130), and obtain

$$\Delta t_{\mathrm{r}} = \frac{1 - (\boldsymbol{n}+\boldsymbol{w}+\boldsymbol{\epsilon})\cdot\boldsymbol{v}_{\mathrm{s}}}{1 - (\boldsymbol{n}+\boldsymbol{w}+\boldsymbol{\epsilon})\cdot\boldsymbol{v}_{\mathrm{r}}}\Delta t_{\mathrm{s}}\,. \tag{134}$$

what generalizes (113). We replace the time differences Δt_{r} and Δt_{s} by the times shown by moving clocks, that is, by the time of the moving system. For that we need (129,130). We get

$$T = \frac{1}{a(v_{\mathrm{r,s}})}\left(t^{(\mathrm{r,s})} - \boldsymbol{\epsilon}^{(\mathrm{r,s})}\cdot\boldsymbol{x}^{(\mathrm{r,s})}\right) \quad \text{and} \quad T = \frac{1}{a(v)}\left(t - \boldsymbol{\epsilon}\cdot\boldsymbol{x}\right). \tag{135}$$

The clocks are at rest in $S_{\mathrm{r,s}}$ and move with $\boldsymbol{u}_{\mathrm{r,s}}$ with respect to $S_{\mathrm{r,s}}$. This implies $\boldsymbol{x}_{\mathrm{r,s}}^{(\mathrm{r,s})} = 0$ and $\boldsymbol{x}_{\mathrm{r,s}} = \boldsymbol{u}_{\mathrm{r,s}}t_{\mathrm{r,s}}$. Therefore we have two relations

$$T_{\mathrm{r,s}} = \frac{1}{a(v_{\mathrm{r,s}})} t_{\mathrm{r,s}}^{(\mathrm{r,s})} \qquad \text{and} \qquad T_{\mathrm{r,s}} = \frac{1}{a(v)} \left(1 - \boldsymbol{\epsilon} \cdot \boldsymbol{u}_{\mathrm{r,s}}\right) t_{\mathrm{r,s}} \tag{136}$$

from which we obtain the measured time differences $\Delta t_{\mathrm{r,s}}^{(\mathrm{r,s})}$ in terms of the coordinate time differences $\Delta t_{\mathrm{r,s}}$

$$\Delta t_{\mathrm{r,s}}^{(\mathrm{r,s})} = \frac{a(v_{\mathrm{r,s}})}{a(v)} \left(1 - \boldsymbol{\epsilon} \cdot \boldsymbol{u}_{\mathrm{r,s}}\right) \Delta t_{\mathrm{r,s}} \,. \tag{137}$$

With the frequencies $\nu_{\mathrm{r,s}} = 1/\Delta t_{\mathrm{r,s}}^{(\mathrm{r,s})}$ we thus obtain from (113)

$$\frac{\nu_{\mathrm{r}}}{\nu_{\mathrm{s}}} = \frac{a(v)}{a(v_{\mathrm{r}})} \frac{a(v_{\mathrm{s}})}{a(v)} \frac{1 - (\boldsymbol{n} + \boldsymbol{v} + \boldsymbol{\epsilon}) \cdot \boldsymbol{v}_{\mathrm{r}}}{1 - (\boldsymbol{n} + \boldsymbol{v} + \boldsymbol{\epsilon}) \cdot \boldsymbol{v}_{\mathrm{s}}} \frac{1 - \boldsymbol{\epsilon} \cdot \boldsymbol{u}_{\mathrm{s}}}{1 - \boldsymbol{\epsilon} \cdot \boldsymbol{u}_{\mathrm{r}}} \,. \tag{138}$$

For a receiver at rest in S ($\boldsymbol{u}_{\mathrm{r}} = 0$) we finally obtain the Doppler formula for arbitrary synchronization

$$\frac{\nu_{\mathrm{r}}}{\nu_{\mathrm{s}}} = \frac{a(v_{\mathrm{s}})}{a(v_{\mathrm{r}})} \frac{1 - \boldsymbol{\epsilon} \cdot \boldsymbol{u}_{\mathrm{s}}}{1 - (\boldsymbol{n} + \boldsymbol{v} + \boldsymbol{\epsilon}) \cdot \boldsymbol{u}_{\mathrm{s}}} \,. \tag{139}$$

The Experiment

Again we calculate the product of the frequencies measured parallel and antiparallel to the velocity of a moving radiating atom

$$\begin{aligned} \frac{\nu_{\mathrm{r}}^{+} \nu_{\mathrm{r}}^{-}}{\nu_{\mathrm{s}}^2} &= \frac{a(v_{\mathrm{s}})}{a(v_{\mathrm{r}})} \frac{1 - \boldsymbol{\epsilon} \cdot \boldsymbol{u}_{\mathrm{s}}}{1 - (\boldsymbol{n} + \boldsymbol{v} + \boldsymbol{\epsilon}) \cdot \boldsymbol{u}_{\mathrm{s}}} \frac{a(v_{\mathrm{s}})}{a(v_{\mathrm{r}})} \frac{1 - \boldsymbol{\epsilon} \cdot \boldsymbol{u}_{\mathrm{s}}}{1 - (-\boldsymbol{n} + \boldsymbol{v} + \boldsymbol{\epsilon}) \cdot \boldsymbol{u}_{\mathrm{s}}} \\ &= \frac{a^2(v_{\mathrm{s}})}{a^2(v_{\mathrm{r}})} \frac{\left(1 - \boldsymbol{\epsilon} \cdot \boldsymbol{u}_{\mathrm{s}}\right)^2}{\left(1 - (\boldsymbol{v} + \boldsymbol{\epsilon}) \cdot \boldsymbol{u}_{\mathrm{s}}\right)^2 - u_{\mathrm{s}}^2} \,. \end{aligned} \tag{140}$$

The result of the experiment is $\nu_{\mathrm{r}}^{+} \nu_{\mathrm{r}}^{-} = \nu_{\mathrm{s}}^2$ so that

$$\frac{a(v_{\mathrm{s}})}{a(v)} \left(1 - \boldsymbol{\epsilon} \cdot \boldsymbol{u}_{\mathrm{s}}\right) = \sqrt{\left(1 - (\boldsymbol{v} + \boldsymbol{\epsilon}) \cdot \boldsymbol{u}_{\mathrm{s}}\right)^2 - u_{\mathrm{s}}^2} \,. \tag{141}$$

This result allows us to determine the factor $a(v')\gamma'/(a(v)\gamma)$ in (132,133)

$$\frac{a(v_{\mathrm{s}})\gamma(v_{\mathrm{s}})}{a(v)\gamma(v)} = \frac{\gamma(v_{\mathrm{s}}) \sqrt{\left(1 - (\boldsymbol{v} + \boldsymbol{\epsilon}) \cdot \boldsymbol{u}_S\right)^2 - u_S^2}}{\gamma(v) \left(1 - \boldsymbol{\epsilon} \cdot \boldsymbol{u}_S\right)} = 1 \,, \tag{142}$$

where we used (69) in order to calculate $\gamma(v_{\mathrm{s}})$ as function of $\gamma(v)$. With this result the transformations (132,133) become the Lorentz transformations between two arbitrarily synchronized reference frames. From the validity of (142) for all v and v_{s} we again infer that $a(v) = a(0)\sqrt{1 - v^2}$. Together with $a(0) = 1$ from (61) we finally get $a(v) = \sqrt{1 - v^2}$. Furthermore, from (101) we obtain $b(v) = 1/\sqrt{1 - v^2}$ and $d(v) = 1$.

As a result, we obtain the Lorentz transformations for arbitrary synchronization [20–25].

7 The Mansouri-Sexl Test Theory

7.1 The Frame

Since most of the experiments are carried through at small velocities, Mansouri and Sexl performed an expansion of the functions $a(v)$,$b(v)$, $d(v)$ and $\boldsymbol{\epsilon}(v)$ with respect to the velocity

$$a(v) = 1 + \left(\alpha - \frac{1}{2}\right) v^2 + \left(\alpha_2 - \frac{1}{8}\right) v^4 + \ldots = 1 + \alpha^{\mathrm{MS}} v^2 + \alpha_2^{\mathrm{MS}} v^4 + \ldots \quad (143)$$

$$b(v) = 1 + \left(\beta + \frac{1}{2}\right) v^2 + \left(\beta_2 + \frac{3}{8}\right) v^4 + \ldots = 1 + \beta^{\mathrm{MS}} v^2 + \beta_2^{\mathrm{MS}} v^4 + \ldots \quad (144)$$

$$d(v) = 1 + \delta v^2 + \delta_2 v^4 + \ldots \quad (145)$$

$$\boldsymbol{\epsilon} \quad = (\epsilon - 1)\boldsymbol{v}\left(1 + \epsilon_2 v^2 + \ldots\right) . \quad (146)$$

The parameter functions are now replaced by a few constant parameters. Here α^{MS} and β^{MS} are parameters originally introduced by Mansouri and Sexl. Our parameters are chosen so that they vanish if SR is valid, compare [26]. In the case of Einstein-synchronization also ϵ vanishes. For simplicity we restrict in the following to first order in v^2. For the next order see [27].

In first order we obtain for the line element

$$\begin{aligned} s^2 = &\left[1 - 2\alpha v^2\right] t^2 - 2\left[\epsilon + (\alpha - \beta - 2\alpha\epsilon - \epsilon_2 + \epsilon\epsilon_2) v^2\right] \boldsymbol{v} \cdot \boldsymbol{x} t \\ &- \left[1 - 2\delta v^2\right] x^2 + \left[\epsilon^2 + 2(\beta - \delta)\right] (\boldsymbol{v} \cdot \boldsymbol{x})^2 , \end{aligned} \quad (147)$$

for the one-way velocity of light

$$c(\vartheta, v) = 1 - \epsilon v \cos\vartheta - \left[\delta - \alpha + (\beta - \delta + \epsilon^2)\cos^2\vartheta\right] v^2 , \quad (148)$$

and for the two-way velocity of light defined in (127)

$$c_{(2)}(\vartheta, v) = 1 + \left[\delta - \alpha + (\beta - \delta)\cos^2\vartheta\right] v^2 . \quad (149)$$

The relative change of the speed of light

$$\frac{\delta_\vartheta c}{c} = (\delta - \beta) v^2 \sin^2\vartheta \quad (150)$$

$$\frac{\delta_v c}{c} = 2\left(\delta - \alpha + (\beta - \delta)\cos^2\vartheta\right) \boldsymbol{v} \cdot \delta\boldsymbol{v} \quad (151)$$

relates this formalism to the model independent description.

With (79) and (125) we can relate the linearized Robertson–parameters to the Mansouri–Sexl–parameters

$$g_2^0 - g_1^0 = \delta - \beta , \qquad g_0^0 - g_1^0 = \beta - \alpha , \qquad g_0^0 = -2\alpha . \quad (152)$$

Since synchronization does not play a role in the interpretation of experiments, the Mansouri–Sexl test theory is equivalent to the linearizes Robertson test theory.

Since we now have only three parameters only three experiments are needed in order to fix the theory.

7.2 Discussion of the Experiments

Isotropy of the Speed of Light

With (148) the phase shift for a general interference experiment is

$$\Delta\phi(\vartheta, v) = \frac{\omega}{c}\{2(l_1 - l_2) + [(2\alpha - \beta - \delta)(l_1 - l_2) + (\delta - \beta)(l_1 + l_2)\cos(2\vartheta)]\, v^2\} \,. \quad (153)$$

For a Michelson–Morley experiment we choose $l_1 = l_2 = l$ and obtain

$$\Delta\phi(\vartheta, v) = 2\frac{l\omega}{c}\left[(\delta - \beta)v^2 \cos(2\vartheta)\right] \,. \quad (154)$$

Independence from the orientation implies

$$\delta - \beta = 0 \,. \quad (155)$$

For resonators we obtain with (26) and (149) the frequency shift

$$\nu(\vartheta, v) = \frac{n}{L}\left\{1 + \left[\delta - \alpha + (\beta - \delta)\cos^2\vartheta\right] v^2\right\} \,, \quad (156)$$

which in the case of isotropy again implies (155). For the comparison of two orthogonally mounted resonators we obtain the relative beat frequency

$$\frac{\nu(\vartheta, v) - \nu(0, v)}{\nu(0, v)} = (\delta - \beta)v^2 \sin^2\vartheta \,. \quad (157)$$

Constancy of the Speed of Light

If we assume isotropy $\delta = \beta$ then we obtain for a Kennedy–Thorndike-experiment, that is $l_1 \neq l_2$ in (153), the phase shift

$$\Delta\phi(v) = 2(l_1 - l_2)\omega\left(1 + (\alpha - \beta)v^2\right) \,. \quad (158)$$

If this does not depend on the velocity of the apparatus, then

$$\alpha - \beta = 0 \,. \quad (159)$$

The frequency in a resonator (156) now is

$$\nu(v) = \frac{cn}{2L}\left[1 + (\beta - \alpha)v^2\right] \,, \quad (160)$$

which does not depend on v if (159) holds.

Time Dilation

Since we discussed these experiments already in the general framework, there is no need to repeat it in this approximation. The result of the easy calculation is is that the exact result $a(v) = \sqrt{1 - v^2}$ in its linearized form now reads $\alpha = 0$. Therefore we obtain from (159) $\beta = 0$ and from (155) $\delta = 0$. Therefore we were able to determine all three parameters α, β and δ from the outcome of three experiments.

8 Discussion

8.1 Summary

We introduces four frames for the discussion of experiments. Three of these frames are special cases of the general framework, see Fig. 10. These special cases are defined by choosing the Einstein synchronization and by a linearization of the theory for small velocities. Physical results should be independent of the chosen synchronization. The description of experiments with arbitrary synchronization forces one to choose appropriate synchronization independent observables.

While this kinematical test theory has the merit for the first time to identify the consequences of certain experiments for the theoretical description, which led to the notion of the "three classical tests" of SR, there are assumptions made which need to be discussed.

8.2 Advantages of Kinematical Test Theories

There are two important and far-reaching advantages:

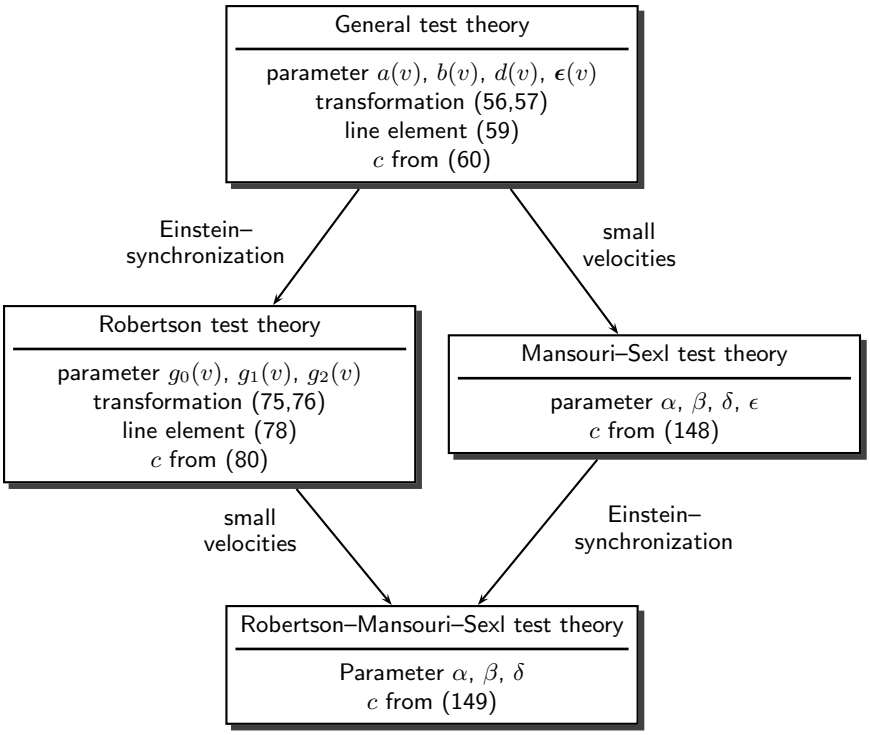


Fig. 10. The relations among the various kinematical test theories. Einstein synchronization connects physically equivalent theories

1. Kinematical test theories are independent from any particle model; they are universal. Since the transformation between frames of reference are under consideration, all physical phenomena are treated in the same way.
2. The test theory is characterized by three parameters only with the consequence that only three experiments are needed in order to fix the theory.

8.3 Disadvantages of Kinematical Test Theories

There are a few severe problems with kinematical test theories based on the assumptions made for setting up this kinematical test theory:

1. The kinematical tests theories need a preferred frame. The choice of a preferred frame may not be unique. Today, one may identify this preferred frame by the cosmological frame defined by the isotropy of the microwave background radiation [28]. Though it is not very probably, it is at least possible in principle, that a stochastic gravitational wave background radiation may define another preferred frame different from the microwave background. Since all estimates describing the degree of validity of SR use the velocity with respect to the preferred frame, the characterization thus depends on our knowledge of cosmology. If we choose another preferred frame, the estimates will change. Therefore, these test theories are intrinsically incomplete. One necessarily needs more input than provided by the kinematical test theory.
2. One assumes a certain geometry of the preferred frame (what has nothing to do with the transformation laws). That means that in Σ one assumes an isotropic speed of light. In principle this also should be subject to experimental proof. One way to handle such a question might be to enlarge the set of parameters by introducing a general propagation through
$$dT^2 = G_1^2 dX^2 - G_2^2 dY^2 - G_3^2 dZ^2 \tag{161}$$
with undetermined parameters G_1, G_2, and G_3. Even more general propagation structures like that of Finslerian structure $dT = f(d\boldsymbol{x})$, f being homogeneous of degree one in $d\boldsymbol{X}$ and non-degenerate, are possible. It should be no problem to carry through the above calculations for this more general setting. However, then more experiments are needed in order to fix the enlarged set parameters. That means, the discussion of experiments for determining the final structure of space-time will be more intriguing.
3. In kinematical test theories the violation of Lorentz invariance can depend on velocity only. A violation of Lorentz invariance may come in through some cosmologically given vector or tensor fields which may occur, e.g., in string theories with spontaneous broken Lorentz symmetry, where the ground state of the space–time geometry shows a broken symmetry which is not present in the formulation of the theory [29,30].
4. Kinematical test theories are not only incomplete, they might be even inconsistent if one considers light to be a consequence of the Maxwell equations. This can be seen as follows: If the light depends on the state of motion of the

laboratory, then also the Maxwell equations have to depend on that state of motion. That means that clocks and rods, which both are heavily determined by the Maxwell equations, also depend on the state of motion. Furthermore, they depend in a material-dependent way on th state of motion. Therefore, there is no unique clock and rod. This, however, is part of the scheme of the kinematical test theories.

5. Kinematical test theories cannot describe birefringence in vacuum which also violates Lorentz invariance.
6. Generalizing this idea, kinematical test theories cannot treat a non–unique c, that is, different limiting velocities for different particles. The dynamics of particles is not treated in these kinematical test theories.
7. Furthermore, it is not possible to describe violations of LI by anisotropic masses or anomalous spin couplings of Dirac particles.
8. Since in our scheme of kinematical test theories we assumed in Σ and, thus, also in other frames S a unique light propagation, it is not possible to describe a hypothetical dependence of c from the velocity of source.

All these problems do not occur in dynamical test theories which, by construction, are complete.

Acknowledgements

I like to thank H. Dittus, D. Giulini, H. Müller, A. Peters, and S. Schiller for helpful discussions. Financial support of the German Space Agency DLR is greatfully acknowledged.

References

1. C.M. Will. The confrontation between general relativity and experiment. *Living Rev. Relativity*, 4: 2001, http://www.livingreviews.org/lrr-2001-4.
2. G. Amelino-Camelia, C. Lämmerzahl, A. Macias, and H. Müller. The search for quantum gravity signals. In A. Macias, C. Lämmerzahl, and D. Nunez, editors, *Gravitation and Cosmology*, p. 30. AIP Conference Proceedings 758, Melville, New York, 2005.
3. D. Mattingly. Modern tests of lorentz invariance. *Living Rev. Relativity*, 8: 2005, http://www.livingreviews.org/lrr-2005-5 (cited on Dec. 15, 2005).
4. C. Lämmerzahl. Relativity and technology. *Ann. Physik (Leipzig)*, 15: 5, 2006.
5. M. Haugan and C.M. Will. Modern tests of special relativity. *Physics Today*, May:69, 1987.
6. M.D. Gabriel and M.P. Haugan. Testing the Einstein Equivalence Principle: Atomic clocks and local Lorentz invariance. *Phys. Rev.*, D 41:2943, 1990.
7. A.P. Lightman and D.L. Lee. Restricted proof that the weak equivalence principle implies the Einstein equivalence principle. *Phys. Rev.*, D 8:364, 1973.
8. H.P. Robertson. Postulate versus observation in the Special Theory of Relativity. *Rev. Mod. Phys.*, 21:378, 1949.

9. R. Mansouri and R.U. Sexl. A test theory of special relativity: I. Simultaneity and clock synchronisation. *Gen. Rel. Grav.*, 8:497, 1977.
10. C.M. Will. *Theory and Experiment in Gravitational Physics (Revised Edition).* Cambridge University Press, Cambridge, 1993.
11. D. Colladay and V.A. Kostelecky. Lorentz–violating extension of the standard model. *Phys. Rev.*, D 58:116002, 1998.
12. C. Lämmerzahl, A. Macias, and H. Müller. Lorentz invariance violation and charge (non-)conservation: A general theoretical frame for extensions of the Maxwell equations. *Phys. Rev.*, D:to appear, 2005.
13. A.A. Michelson and E.W. Morley. On the relative motion of the Earth and the luminiferous ether. *Am. J. Sci.*, 34:333, 1887.
14. H. Müller, S. Herrmann, A. Saenz, A. Peters, and C. Lämmerzahl. Optical cavity tests of Lorentz invariance for the electron. *Phys. Rev.*, D 68:116006, 2003.
15. R.S. Shankland, S.W. McCuskey, F.C. Leone, and G. Kuerti. New analysis of the interferometer observations of Dayton C. Miller. *Rev. Mod. Phys.*, 27:167, 1955.
16. E.W. Morley and D.C. Miller. (letter to Lord Kelvin). *Phil. Mag.*, 8:753, 1904.
17. E.W. Morley and D.C. Miller. Report of an experiment to detect the FitzGerald–Lorentz–effect. *Phil. Mag.*, 9:680, 1905.
18. L. Essen. A new aetherdrift experiment. *Nature*, 175:793, 1955.
19. V.A. Fock. *Theorie von Raum, Zeit und Gravitation.* Akademie–Verlag, Berlin, 1960.
20. H. Reichenbach. *Axiomatik der Raum–Zeit Lehre.* deGruyter, Berlin, 1924.
21. A. Grünbaum. *Philosophical Problems of Space and Time.* D. Reidel, Dordrecht, 1973.
22. J.A. Winnie. Special Relativity without one-way velocity assumptions: Part I. *Philosophy of Science*, 37:81, 1968.
23. W.F. Edwards. Special relativity in aniosotropic space. *Am. J. Phys.*, 31:482, 1963.
24. J.A. Winnie. Special Relativity without one-way velocity assumptions: Part II. *Philosophy of Science*, 37:223, 1968.
25. Y.Z. Zhang. Test theories of special relativity. *Gen. Rel. Grav.*, 27:475, 1994.
26. C.M. Will. The confrontation between general relativity and experiment: a 1992 update. *Int. J. Mod. Phys.*, D 1:13, 1992.
27. C. Lämmerzahl, C. Braxmaier, H.-J. Dittus, H. Müller, A. Peters, and S. Schiller. Kinematical test theories for special relativity: A comparison. *Int. J. Mod. Phys.*, D 11:1109, 2002.
28. C.H. Lineweaver, L. Tenorio, G.F. Smoot, P. Keegstra, A.J. Banday, and P. Lubin. The dipole observed in the cobe dmr 4 year data. *Astroph. J.*, 470:38, 1996.
29. V.A. Kostelecky and S. Samuel. Spontaneous breaking of Lorentz symmetry in string theory. *Phys. Rev.*, D 39:1611, 1989.
30. R. Bluhm. Overview of the SME: Implications and phenomenology of Lorentz violation. Lect. Notes Phys. **702**, pp. 191–226, 2006.

Test of Lorentz Invariance Using a Continuously Rotating Optical Resonator

S. Herrmann[1], A. Senger[1], E. Kovalchuk[1,2], H. Müller[1,2,3], and A. Peters[1]

[1] Institut für Physik, Humboldt Universität zu Berlin, 10117 Berlin, Germany
sven.herrmann@physik.hu-berlin.de,
[2] Fritz–Haber–Institut der Max–Planck–Gesellschaft, Faradayweg 4-6, 14195 Berlin, Germany
[3] Physics Department, Stanford University, Stanford, CA 94305, USA

1 Introduction

Local Lorentz Invariance (LLI), stating that locally physical laws are identical in all inertial reference frames, constitutes the basis of special relativity and is an essential ingredient of both the standard model of particle physics and the theory of general relativity. A well known test experiment for this fundamental symmetry is the Michelson-Morley (MM) experiment (Fig. 1), which even predated the formulation of special relativity. First performed by A.A. Michelson in Potsdam in 1881 it was later repeated at increased precision together with E.W. Morley in Cleveland, Ohio, in 1887 [1]. While their motivation was to reveal an anisotropy of the speed of light c due to Earth's motion through an ether medium, that had been postulated as a carrier for electromagnetic waves, they were left with an unexpected null result. This was only clearly understood when Einstein formulated the theory of special relativity in 1905 building on the constancy of c, i.e. its independence on laboratory velocity and orientation. The latter has since been verified experimentally at improved precision by numerous repetitions of the MM-experiment (Fig. 2), providing a firm experimental basis for special relativity so far.

Today however, tiny violations of fundamental principles such as LLI are discussed within several attempts aiming to formulate a unifying theory of quantum gravity. Modern high precision test experiments are thus considered as important contributions to these attempts, as they might either rule out or possibly reveal the presence of such effects at some level of measurement precision. The consequences of broken Lorentz symmetry have only recently been described within a very general and consistent framework called Standard Model Extension (SME) [2]. This test model adds to the Lagrangian of the standard model of particle physics all LLI-violating terms that can be formed from the known fields and Lorentz tensors. In particular it models an anisotropy of the speed of light in the most general way, thus providing the possibility for a comprehensive and consistent analysis of modern descendants of the MM-experiment. Such state of

S. Herrmann et al.: *Test of Lorentz Invariance Using a Continuously Rotating Optical Resonator*, Lect. Notes Phys. **702**, 385–400 (2006)
DOI 10.1007/3-540-34523-X_13
© Springer-Verlag Berlin Heidelberg 2006

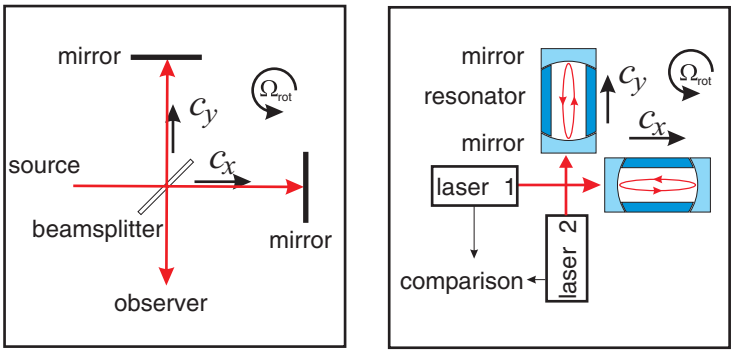


Fig. 1. Schematic of Michelson-Morley experiments. *Left*: classic Michelson interferometer. *Right*: Modern version comparing the frequencies of two lasers stabilized to resonances of high finesse cavities ($\nu_{\mathrm{res}} \sim c/L$)

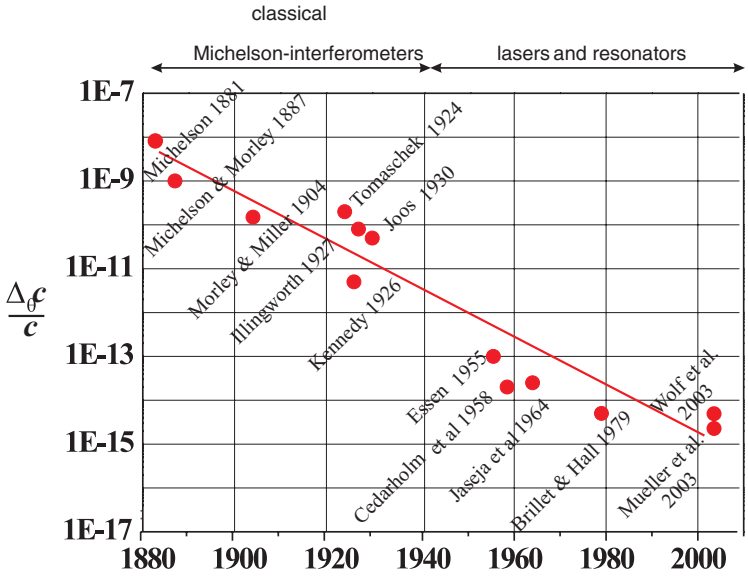


Fig. 2. Improvements of the Michelson-Morley experiment since 1881

the art MM-experiments employ high finesse electromagnetic resonators, whose eigenfrequencies depend on the speed of light c in a geometry dependent way ($\nu \sim c/L$ for a linear optical Fabry-Perot cavity of length L). Thus a measurement of the eigenfrequency of a resonator as its orientation is varied, should reveal an anisotropy of c/L.

During recent years modern versions of the MM-experiment have been realized using optical [3, 4] as well as microwave resonators [5–7]. So far these experiments relied solely on Earth's rotation for varying resonator orientation, which was only possible using cryogenically cooled resonators that exhibit very low

frequency drifts. However, actively rotating the setup as done in a classic experiment by Brillet and Hall [8] offers two strong benefits: (i) the rotation rate can be matched to the timescale of optimal resonator frequency stability and (ii) the statistics can be significantly improved by performing thousands of rotations per day. Applying state of the art technology as done in the modern non rotating experiments, these advantages should allow for tests improved by orders of magnitude – assuming that systematic effects induced by the active rotation can be kept sufficiently low. Following this approach we have set up such an actively rotating MM-experiment, and here we report on first results that already yield improved limits on LLI-violation. Similar concurrent experiments are currently performed either using continuously rotating microwave resonators [9] or cryogenic optical resonators [10], whose orientation is varied by 90°.

2 Setup

The basic ingredient of the experiment is an optical cavity which is continuously rotated on a precision air bearing turntable. The cavity is fabricated from fused silica, it has a length $L = 3\,\mathrm{cm}$ and linewidth 20 kHz. Its resonance frequency is compared to that of a stationary cavity oriented north-south (L = 10 cm, 10 kHz linewidth) [11]. Each cavity is mounted inside a thermally shielded vacuum chamber. The cavity resonance frequencies are interrogated by two diode pumped Nd:YAG lasers (1064 nm), coupled to the cavities through windows in the vacuum chambers, and stabilized using the Pound-Drever-Hall method [12]. The laser frequency is tuned by laser crystal temperature and laser crystal strain using a piezo electric crystal.

The table rotation rate $\omega_{\mathrm{rot}} = 2\pi/T$ is set to $T \sim 43\,\mathrm{s}$ (~ 2000 rotations/day) matching the timescale of optimum cavity stability. The relative root Allan variance at this integration time reaches $\Delta\nu/\nu = 1.2 \times 10^{-14}$ as shown in Fig. 4. At this rotation rate it is also possible to rely on the excellent thermal isolation properties of the vacuum chambers at room temperature (time constant $\sim 10\,\mathrm{h}$). The residual temperature drift of the resonance frequencies is on the order of 1 MHz/day, which is comparatively high but sufficiently linear to be cleanly separated from a potential LLI-violation signal at $2\omega_{\mathrm{rot}}$. Figure 3 gives a schematic view of the rotating setup. Electrical connections are made via an electric 15 contact slip ring assembly on top. To measure the frequency difference $\Delta\nu$ of both lasers, a fraction of the rotating laser's light leaves the table aligned with the rotation axis and is then overlapped with light from the stationary laser on a high speed photodetector. The resulting beat note at the difference frequency $\Delta\nu \sim 2\,\mathrm{GHz}$ is read out at a sampling rate of 1/s after down conversion to about 100 MHz. Amplitude modulation arising from periodic beam displacement relative to the detector due to table rotation could be reduced below 10% and does not affect the outcome of the frequency measurement.

Substantial effort was spent on minimizing systematic effects associated with turntable rotation (see Fig. 5). Besides good thermal and electromagnetic shielding, prevention of cavity deformations due to external forces is most importantly

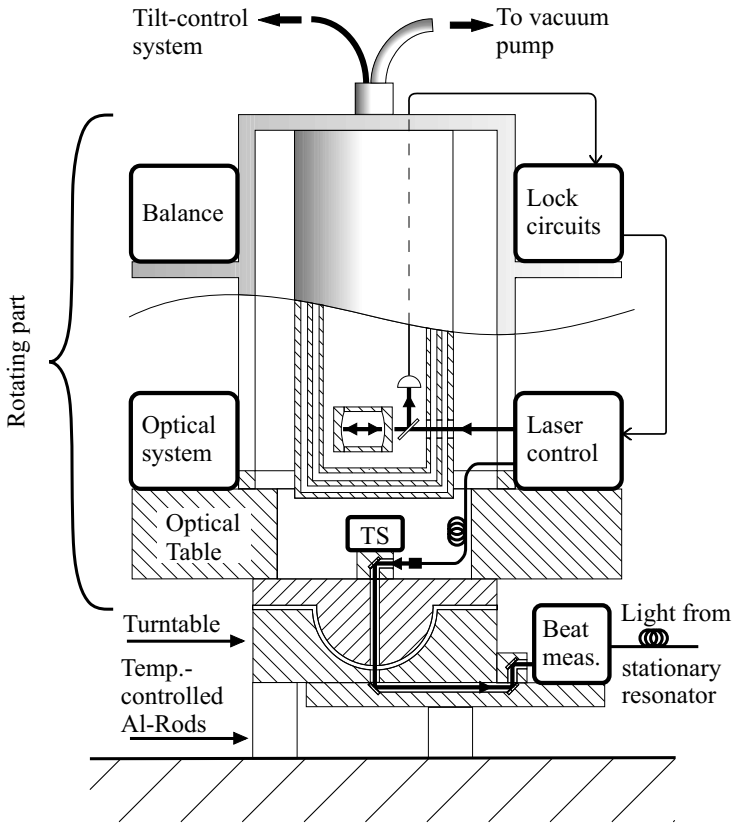


Fig. 3. Setup of the rotating part of the experiment. A high performance turntable is applied specified for rotation axis wobble <1μrad. The center-of-mass of the setup is carefully balanced and tilt is monitored using an electronic bubble level tilt sensor (TS)

involved here. Such forces are either of centrifugal or gravitational origin. If not supported in a perfectly symmetric manner the latter causes a deformation when tilted against the horizontal. The observed relative frequency change for our setup is $1.5 \times 10^{-16}/\mu\text{rad}$. Tilts which vary as a function of the orientation of the turntable enter the analysis of the experiment as a systematic error, and have to be suppressed by keeping the rotation axis vertical and preventing wobble in the setup. The latter is achieved by employing a turntable with intrinsic wobble $< 1\,\mu\text{rad}$ and carefully balancing the center-of-mass of the rotating part. To prevent long term variations of rotation axis tilt, an active tilt control is applied. Similar to the scheme described in [13], we place the table on three aluminum cylinders, 20 cm in length, two of which can be heated independently in order to use thermal expansion ($5\,\mu\text{m}/^\circ\text{C}$) to compensate slow tilt variations. The tilt is monitored using an electronic bubble level sensor of 0.1 μrad resolution placed at the turntable center. A computer program transforms the rotating sensor's read-

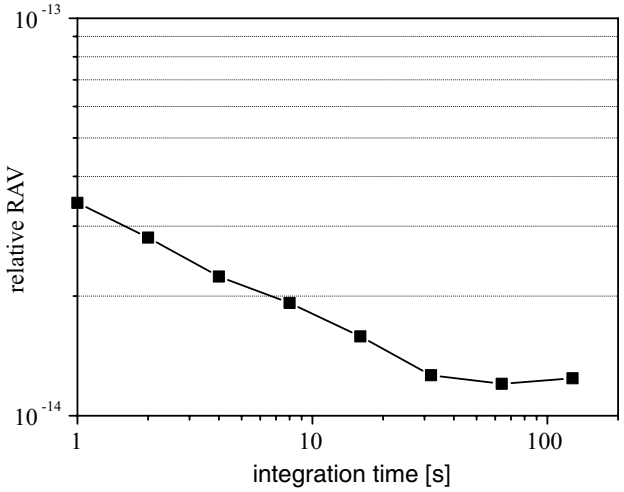


Fig. 4. Laser lock stability at different integration times expressed by relative root Allan variance. After removal of a linear drift the optimum frequency stability is obtained at about 50 s integration time. The rotation rate is chosen slightly below in favor of improved statistics due to an increased number of rotations per day

out to a non rotating reference frame and controls the heating of the aluminum feet, which is done in a closed loop. Typical tilt variations of the laboratory's ground floor are several 10 μrad/day that would give rise to (varying) systematic effects at $2\omega_{\rm rot}$ of up to one part in 10^{-14} without tilt control. The active stabilization reduces tilt variations to $< 1\,\mu$rad corresponding to systematic tilt induced effects $< 10^{-16}$.

To prevent systematic effects arising from modulated centrifugal forces the rotation rate is phase locked to a stable reference oscillator. Applying a high resolution encoder implemented in the turntable allowed us to finally suppress deviations from ideal rotation to less than 0.1° at any instant. Our estimates on the influence of varying centrifugal forces on resonator stability show that this is sufficient at the current level of sensitivity but might become an issue when aiming for the 10^{-17} level of measurement precision.

3 LLI-Violation Signal According to SME

For the photonic sector of the SME the extended Lagrangian takes the form

$$L = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}(k_{AF})^{\kappa}\epsilon_{\kappa\lambda\mu\nu}A^{\lambda}F^{\mu\nu} - \frac{1}{4}(k_F)_{\kappa\lambda\mu\nu}F^{\kappa\lambda}F^{\mu\nu}\,, \qquad (1)$$

where $F_{\mu\nu}$ is the electromagnetic field tensor and A^{μ} the vector potential. The first term is the ordinary Maxwell Lagrangian and the two additional terms are the LLI-violating extensions. The second term is expected to vanish for

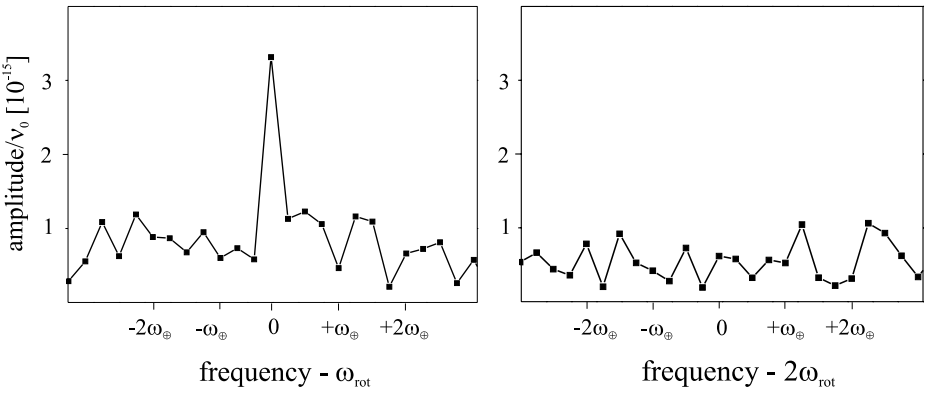


Fig. 5. Fourier transform of a 4-day data set starting on February 18, 2005, with active tilt control applied. *Left*: A small residual effect is present at $\omega_{\rm rot}$ *Right*: No peak is visible at $2\omega_{\rm rot}$ nor at the siderial sidebands

theoretical reasons and is constrained experimentally to levels well below those relevant here [14]. The tensor (k_F) within the third term contains 19 independent parameters, which can be arranged into one scalar κ_{tr}, and four traceless 3×3 matrices: $\tilde{\kappa}_{e-}$, $\tilde{\kappa}_{o+}$, $\tilde{\kappa}_{e+}$ and $\tilde{\kappa}_{o-}$. While κ_{tr} is related to the one way speed of light [15], the elements of the latter two matrices are restricted to values $< 2\times10^{-32}$ by astrophysical observations [16]. The remaining matrices $\tilde{\kappa}_{e-}$ and $\tilde{\kappa}_{o+}$ contain 8 parameters that describe a boost dependent ($\tilde{\kappa}_{o+}$, antisymmetric) and a boost independent ($\tilde{\kappa}_{e-}$, symmetric) anisotropy of the speed of light. Recent measurements have restricted 7 of these elements to a level of 10^{-11} respectively 10^{-15} [3,5–7]. The remaining component $\tilde{\kappa}_{e-}^{ZZ}$ can only be determined in actively rotating experiments, thus it was not accessible in these experiments relying solely on Earth's rotation.

The implications of (1) on the eigenfrequency of a linear Fabry-Perot resonator stationary in an Earth based laboratory have been elaborated in detail in [2]. The convention there is to refer all values of SME-parameters to a Sun centered celestial equatorial frame (SCCEF) that has the X-axis pointing towards vernal equinox, the Z-axis pointing towards the celestial north pole and the Y-axis is chosen accordingly to complete the right handed dreibein. The coordinate axes x, y, z for the earth based laboratory (on the northern hemisphere) are commonly chosen such that the x-axis points south, the y-axis points east and the z-axis points vertically upwards. Transition from the SCCEF to this laboratory frame is done using the transformations given in [2]. They involve Lorentz transformations according to Earth's orbital and rotational boost ($\beta_\oplus \sim 10^{-4}$, $\beta_L \sim 10^{-6}$ and a rotation matrix

$$R = \begin{pmatrix} \cos\chi\cos\omega_\oplus T_\oplus & \cos\chi\sin\omega_\oplus T_\oplus & -\sin\chi \\ -\sin\omega_\oplus T_\oplus & \cos\omega_\oplus T_\oplus & 0 \\ \sin\chi\cos\omega_\oplus T_\oplus & \sin\chi\sin\omega_\oplus T_\oplus & \cos\chi \end{pmatrix} . \qquad (2)$$

$\omega_\oplus$ is the frequency of a siderial day and the time axis $T_\oplus$ is fixed by $T_\oplus = 0$ when the laboratory y axis and the SCCEF Y-axis coincide. As shown in [2] the resonance frequency change $\delta\nu$ induced by the LLI-violating extension within (1) for a linear optical cavity stationary in the laboratory frame is

$$\frac{\delta\nu}{\nu_0} = A(t) + B(t)\sin 2\theta + C(t)\cos 2\theta \,, \tag{3}$$

where ν_0 is the undisturbed laser frequency (2.82×10^{14} Hz for our setup) and θ is the angle relative to the laboratory x-axis. The amplitudes $A(t), B(t)$ and $C(t)$ depend on elements of the $\tilde{\kappa}$-matrices (stated in the SCCEF) and exhibit a modulation due to Earth's rotation and orbital motion as shown in [2]. For a rotating cavity we adopt $\theta = \omega_{\rm rot}T$ with $T = 0$ when the resonator axis and the laboratory x-axis coincide. The frequency difference between a rotating and a stationary cavity oriented north south ($\theta = 0$) then is

$$\frac{\Delta\nu}{\nu_0} = \frac{\delta\nu_{\rm r} - \delta\nu_{\rm s}}{\nu_0} = B(t)\sin 2\omega_{\rm rot}T + C(t)\cos 2\omega_{\rm rot}T - C(t) \,, \tag{4}$$

where $B(t)$ and $C(t)$ vary according to

$$\begin{aligned} B(t) = {} & B_{\rm dc} + B_{s1}\sin\omega_\oplus T_\oplus + B_{c1}\cos\omega_\oplus T_\oplus \\ & + B_{s2}\sin 2\omega_\oplus T_\oplus + B_{c2}\cos 2\omega_\oplus T_\oplus, \end{aligned} \tag{5}$$

$$\begin{aligned} C(t) = {} & C_{\rm dc} + C_{s1}\sin\omega_\oplus T_\oplus + C_{c1}\cos\omega_\oplus T_\oplus \\ & + C_{s2}\sin 2\omega_\oplus T_\oplus + C_{c2}\cos 2\omega_\oplus T_\oplus \,. \end{aligned} \tag{6}$$

The explicit dependence of the Fourier coefficients $B_{\rm ab}$ and $C_{\rm ab}$ [17] on the SME-parameters is given in Table 1 according to [2]. As the experiment is designed for optimum frequency stability at a timescale of several rotations ($\omega_{\rm rot} \gg \omega_\oplus$) the slow variations due to the last term $C(t)$ in (4) are not relevant for the analysis here. Note further that the same linear combinations of SME-parameters are involved in the $B_{\rm ab}$ and $C_{\rm ab}$ coefficients so there are only 5 independent combinations for the 8 SME-parameters. While the siderial modulation of $\tilde{\kappa}_{o+}$-terms within these linear combinations can in principle be used to determine limits on individual SME-parameters, this is only possible for a measurement spanning > 1 year.

An alternative representation of the LLI-violation signal can be found, if (5) and (6) are inserted into (4). Expansion of the the resulting expression then yields 2×5 terms oscillating at $2\omega_{\rm rot}$ and the sidebands at $2\omega_{\rm rot} \pm \omega_\oplus$ and $2\omega_{\rm rot} \pm 2\omega_\oplus$:

$$\frac{\Delta\nu}{\nu_0} = \sum_{k=-2}^{2} \left(B_k \sin[(2\omega_{\rm rot} + k\omega_\oplus)T_\oplus - 2\phi] + C_k \cos[(2\omega_{\rm rot} + k\omega_\oplus)T_\oplus - 2\phi]\right) \,, \tag{7}$$

where the constant phase shift ϕ accounts for the difference in time axes T and $T_\oplus$ as defined above. The dependence of the sideband amplitudes B_k and

Table 1. Fourier coefficients $B_{\rm ab}$ and $C_{\rm ab}$ of (5) and (6) related to photonic SME-parameters according to [2]. Relations are stated to first order in orbital boost. $\beta_\oplus = 10^{-4}$ is the boost parameter, $\chi = 37^\circ$ is the colatitude of the Berlin laboratory and $\eta = 23^\circ$ is the tilt of Earth's axis within the SCCEF. $\Phi = \Omega_\oplus T'$ gives the siderial phase relative to $T' = 0$ when the Earth passes vernal equinox. $\Omega_\oplus$ denotes the frequency of a siderial year

$C_{\rm ab}/B_{\rm ab}$	SME-parameters
$C_{\rm dc}$	$\frac{3}{8}\sin^2\chi\kappa_{e-}^{ZZ}$ $-\frac{1}{4}\sin^2\chi[\cos\eta\beta_\oplus\kappa_{o+}^{XZ}\cos\phi + 2\sin\eta\beta_\oplus\kappa_{o+}^{XY}\cos\phi + \beta_\oplus\kappa_{o+}^{YZ}\sin\phi]$
$C_{\rm s1}$	$-\frac{1}{2}\cos\chi\sin\chi\kappa_{e-}^{YZ}$ $+\frac{1}{2}\cos\chi\sin\chi[\cos\eta\beta_\oplus\kappa_{o+}^{XY}\cos\phi - \sin\eta\beta_\oplus\kappa_{o+}^{XZ}\cos\phi]$
$C_{\rm c1}$	$-\frac{1}{2}\cos\chi\sin\chi\kappa_{e-}^{XZ}$ $+\frac{1}{2}\cos\chi\sin\chi[\sin\eta\beta_\oplus\kappa_{o+}^{YZ}\cos\phi - \beta_\oplus\kappa_{o+}^{XY}\sin\phi]$
$C_{\rm s2}$	$\frac{1}{4}(1+\cos^2\chi)\kappa_{e-}^{XY}$ $-\frac{1}{4}(1+\cos^2\chi)[\cos\eta\beta_\oplus\kappa_{o+}^{YZ}\cos\phi + \beta_\oplus\kappa_{o+}^{XZ}\sin\phi]$
$C_{\rm c2}$	$\frac{1}{8}(1+\cos^2\chi)[\kappa_{e-}^{XX} - \kappa_{e-}^{YY}]$ $-\frac{1}{4}(1+\cos^2\chi)[\cos\eta\beta_\oplus\kappa_{o+}^{XZ}\cos\phi - \beta_\oplus\kappa_{o+}^{YZ}\sin\phi]$
$B_{\rm dc}$	0
$B_{\rm s1}$	$-C_{\rm c1}/\cos\chi$
$B_{\rm c1}$	$C_{\rm s1}/\cos\chi$
$B_{\rm s2}$	$-2C_{\rm c2}\cos\chi/(1+\cos^2\chi)$
$B_{\rm c2}$	$2C_{\rm s2}\cos\chi/(1+\cos^2\chi)$

C_k on the SME-parameters is given in Table 2. The sideband amplitudes can be transformed to the Fourier coefficients of Table 1 using the simple relations given in Table 3.

While for the analysis of this and other recent MM-experiments the consequences of LLI-violation in the photonic sector of the SME have been considered as modelled in [2], we note that the SME allows for an analysis of these experiments including further consequences of LLI-violation which have not been treated in [2]. In [18] we have investigated the influence of $(k_F)_{\kappa\lambda\mu\nu}$ on the cavity length, arguing that the Lorentz-violating electrodynamics necessarily leads to a modification of the Coulomb interaction, which in turn results in a change of the cavity length. This length is also modified as a result from LLI-violation in the electronic sector of the SME, the implications of which we have elaborated

Table 2. Siderial sideband amplitudes C_k and B_k according to (7) related to SME-parameters. See Table 1 for further explanations

C_k/B_k	SME-parameters
C_0	$\frac{3}{8}\sin^2\chi\tilde{\kappa}_{e-}^{ZZ}$ $-\frac{1}{4}\beta_\oplus\sin^2\chi[(\cos\eta\tilde{\kappa}_{o+}^{XZ}+2\sin\eta\tilde{\kappa}_{o+}^{XY})\cos\Phi-\tilde{\kappa}_{o+}^{YZ}\sin\Phi]$
C_{+1}	$-\frac{1}{4}(1+\cos\chi)\sin\chi\tilde{\kappa}_{e-}^{XZ}$ $+\frac{1}{4}\beta_\oplus(1+\cos\chi)\sin\chi[\sin\eta\tilde{\kappa}_{o+}^{YZ}\cos\Phi-\tilde{\kappa}_{o+}^{XY}\sin\Phi]$
C_{+2}	$\frac{1}{16}(1+\cos\chi)^2[\tilde{\kappa}_{e-}^{XX}-\tilde{\kappa}_{e-}^{YY}]$ $-\frac{1}{8}\beta_\oplus(1+\cos\chi)^2[\cos\eta\tilde{\kappa}_{o+}^{XZ}\cos\Phi-\tilde{\kappa}_{o+}^{YZ}\sin\Phi]$
C_{-1}	$C_{+1}\dfrac{(\cos\chi-1)}{(\cos\chi+1)}$
C_{-2}	$C_{+2}\dfrac{(\cos\chi-1)^2}{(\cos\chi+1)^2}$
B_0	0
B_{+1}	$-\frac{1}{4}(1+\cos\chi)\sin\chi\tilde{\kappa}_{e-}^{YZ}$ $\frac{1}{4}(1+\cos\chi)\sin\chi\beta_\oplus[\cos\eta\tilde{\kappa}_{o+}^{XY}\cos\phi-\sin\eta\tilde{\kappa}_{o+}^{XZ}\cos\phi]$
B_{+2}	$\frac{1}{8}(1+\cos\chi)^2\tilde{\kappa}_{e-}^{XY}$ $-\frac{1}{8}\beta_\oplus(1+\cos\chi)^2[\cos\eta\tilde{\kappa}_{o+}^{YZ}\cos\Phi+\tilde{\kappa}_{o+}^{XZ}\sin\Phi]$
B_{-1}	$-B_{+1}\dfrac{(\cos\chi-1)}{(\cos\chi+1)}$
B_{-2}	$-B_{+2}\dfrac{(\cos\chi-1)^2}{(\cos\chi+1)^2}$

Table 3. Relation of sideband amplitudes B_k, C_k of (7) to Fourier coefficients $B_{\mathrm{ab}}, C_{\mathrm{ab}}$ of (5) and (6)

C_{ab}	C_k	B_{ab}	B_k
C_{dc}	C_0	B_{dc}	B_0
C_{s1}	$B_{+1}-B_{-1}$	B_{s1}	$C_{-1}-C_{+1}$
C_{c1}	$C_{+1}+C_{-1}$	B_{c1}	$B_{-1}+B_{+1}$
C_{s2}	$B_{+2}-B_{-2}$	B_{s2}	$C_{-2}-C_{+2}$
C_{c2}	$C_{+2}+C_{-2}$	B_{c2}	$B_{-2}+B_{+2}$

in [19]. A complete analysis of this experiment with respect to these further aspects of LLI-violation is yet to be done.

4 LLI-Violation Signal According to RMS

The MM-experiment is also often interpreted according to a kinematical test theory, formulated by Robertson [21] and Mansouri and Sexl [22] (RMS). This test theory assumes a preferred isotropic frame Σ, commonly adopted to be the cosmic microwave background (CMB). Within generalized transformations between Σ and a laboratory frame S moving at velocity v relative to Σ three test parameters (α, β, δ) are introduced. If Lorentz Invariance is valid these take the values $\alpha = -\frac{1}{2}, \quad \beta = \frac{1}{2}$ and $\delta = 0$ establishing the Lorentz transformations between Σ and S in their familiar form. Deviations from these values lead to modifications in time dilation and to an anisotropy as well as a boost dependence of c within S. Anisotropy of $c = c(\theta)$ is described according to

$$\Delta c(\theta)/c = \mathcal{B}\frac{v^2}{c^2}\sin^2\theta(t)\ , \tag{8}$$

where $\mathcal{B}$ abbreviates the parameter combination $(\beta - \delta - \frac{1}{2})$. θ is the angle between the direction of light propagation and v. From the non rotating MM-experiment described in [3] $\mathcal{B}$ has been restricted to $\mathcal{B} = (2.2 \pm 1.5) \times 10^{-9}$ [23].

From (8) the relative frequency change between two resonators oriented relative to v at angles $\theta_1(t)$ and $\theta_2(t)$ can be obtained according to

$$\frac{\Delta(\nu_1 - \nu_2)}{\nu_o} = \mathcal{B}\frac{v^2}{c^2}\left[\sin^2\theta_1(t) - \sin^2\theta_2(t)\right]\ . \tag{9}$$

To determine the explicit time dependence of this expression we refer to the SCCEF as defined above and evaluate the relation

$$\sin^2\theta_i(t) = 1 - \left(\frac{\mathbf{v}(t)\mathbf{e}_i(t)}{v^2}\right)^2\ , \tag{10}$$

where $\mathbf{e}_1(t)$ and $\mathbf{e}_2(t)$ are the unit vectors along the axis of the stationary and the rotating resonator. In the laboratory frame these are

$$(\mathbf{e}_1)_{\text{lab}} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad , \quad (\mathbf{e}_2)_{\text{lab}} = \begin{pmatrix} \cos 2\omega_{\text{rot}}T \\ \sin 2\omega_{\text{rot}}T \\ 0 \end{pmatrix}\ . \tag{11}$$

Rotation to the SCCEF then results in

$$\mathbf{e}_1(t) = \begin{pmatrix} \cos\chi\cos\omega_\oplus T_\oplus \\ \cos\chi\sin\omega_\oplus T_\oplus \\ -\sin\chi \end{pmatrix}\ , \tag{12}$$

$$\mathbf{e}_2(t) = \begin{pmatrix} \cos\chi\cos\omega_\oplus T_\oplus \cos(\omega_{\rm rot}T_\oplus+\phi) - \sin\omega_\oplus T_\oplus \sin(\omega_{\rm rot}T_\oplus+\phi) \\ \cos\chi\cos(\omega_{\rm rot}T_\oplus+\phi)\sin\omega_\oplus T_\oplus + \cos\omega_\oplus T_\oplus \sin(\omega_{\rm rot}T_\oplus+\phi) \\ -\cos(\omega_{\rm rot}T_\oplus+\phi)\sin\chi \end{pmatrix} , \quad (13)$$

where the phase ϕ accounts for the difference in timescales of T and $T_\oplus$ as defined in section 3 and v is the velocity of the laboratory relative to the CMB. If we only consider the constant leading term of v neglecting Earth's orbital and rotational boosts, its orientation in the SCCEF is given by

$$\mathbf{v}(t) = v \begin{pmatrix} \cos\alpha\cos\beta \\ \sin\alpha\cos\beta \\ -\sin\beta \end{pmatrix} , \quad (14)$$

with $\alpha = 168°$, $\beta = -6°$ and $v = 370\,\mathrm{km/s}$. An evaluation of (9) using the above expressions finally yields a signal oscillating at $2\omega_{\rm rot}$ and the sidebands at $\omega_\oplus$ and $2\omega_\oplus$ which can be expressed exactly as done in (7), where the sideband amplitudes B_k and C_k are now related to the RMS-parameter $\mathcal{B}$ as shown in Table 4.

Again we can find an alternative description of this signal by rearranging it as a signal oscillating at $2\omega_{\rm rot}$ similar to expression (4), where the amplitudes $C(t)$ and $B(t)$ are modulated analogous to (5) and (6). The Fourier coefficients $B_{\rm ab}$ and $C_{\rm ab}$ then are related to $\mathcal{B}$ as stated in Table 5.

Table 4. Siderial sideband amplitudes C_k and B_k analogous to (7) related to RMS-parameter $\mathcal{B}$. v is the velocity of the laboratory relative to the CMB (neglecting Earth's orbital and rotational boost here). $\alpha = 168°$ and $\gamma = -6°$ fix the orientation of v in the SCCEF

C_k/B_k	RMS-parameter $\mathcal{B}$
C_0	$\frac{1}{8}(-1+3\cos 2\gamma)\sin^2\chi\mathcal{B}\frac{v^2}{c^2}$
C_{+1}	$-\frac{1}{4}\cos\alpha\sin 2\gamma(\cos\chi+1)\sin\chi\mathcal{B}\frac{v^2}{c^2}$
C_{+2}	$-\frac{1}{8}\cos 2\alpha\cos^2\gamma(1+\cos\chi)^2\mathcal{B}\frac{v^2}{c^2}$
C_{-1}	$C_{+1}\dfrac{\cos\chi-1}{\cos\chi+1}$
C_{-2}	$C_{+2}\dfrac{(1-\cos\chi)^2}{(1+\cos\chi)^2}$
B_0	0
B_{+1}	$-\frac{1}{4}\sin\alpha\sin 2\gamma(1+\cos\chi)\sin\chi\mathcal{B}\frac{v^2}{c^2}$
B_{+2}	$-\frac{1}{8}\sin 2\alpha\cos^2\gamma(1+\cos\chi)^2\mathcal{B}\frac{v^2}{c^2}$
B_{-1}	$-B_{+1}\dfrac{\cos\chi-1}{\cos\chi+1}$
B_{-2}	$B_{+2}\dfrac{(1-\cos\chi)^2}{(1+\cos\chi)^2}$

Table 5. Fourier coefficients $C_{\rm ab}$ and $B_{\rm ab}$ analogous to (5) and (6) related to the RMS-parameter $\mathcal{B}$. See Table 4 for further explanations

$C_{\rm ab}/B_{\rm ab}$	RMS-parameter $\mathcal{B}$
C_{dc}	$\frac{1}{8}(-1+3\cos 2\gamma)\sin^2\chi\mathcal{B}\frac{v^2}{c^2}$
$C_{\rm s1}$	$-\frac{1}{4}\sin\alpha\sin 2\gamma\sin 2\chi\mathcal{B}\frac{v^2}{c^2}$
$C_{\rm c1}$	$-\frac{1}{4}\cos\alpha\sin 2\gamma\sin 2\chi\mathcal{B}\frac{v^2}{c^2}$
$C_{\rm s2}$	$-\frac{1}{4}\sin 2\alpha\cos^2\gamma(1+\cos^2\chi)\mathcal{B}\frac{v^2}{c^2}$
$C_{\rm c2}$	$-\frac{1}{4}\cos 2\alpha\cos^2\gamma(1+\cos^2\chi)\mathcal{B}\frac{v^2}{c^2}$
$B_{\rm dc}$	0
$B_{\rm s1}$	$-C_{\rm c1}/\cos\chi$
$B_{\rm c1}$	$C_{\rm s1}/\cos\chi$
$B_{\rm s2}$	$-2C_{\rm c2}\cos\chi/(1+\cos^2\chi)$
$B_{\rm c2}$	$2C_{\rm s2}\cos\chi/(1+\cos^2\chi)$

5 Data Analysis

Here we present an analysis of first data collected during 15 measurement runs of 24 h to 100 h in length taken between December 2004 and March 2005. The data comprises 29 days in total, including about 70000 table rotations.

For the SME-analysis we can choose one of the two equivalent representations of the LLI-violation signal described above. Here we will follow the approach that determines the siderial time dependence of $C(t)$ and $B(t)$ in (4) encoded in the 2×5 Fourier coefficients $B_{\rm ab}$ and $C_{\rm ab}$ of (5) and (6). The procedure is as follows: We sample the data from each measurement run into subsets of 10 table rotations each (see Fig. 6) and to each subset we fit (4) extended in the following way :

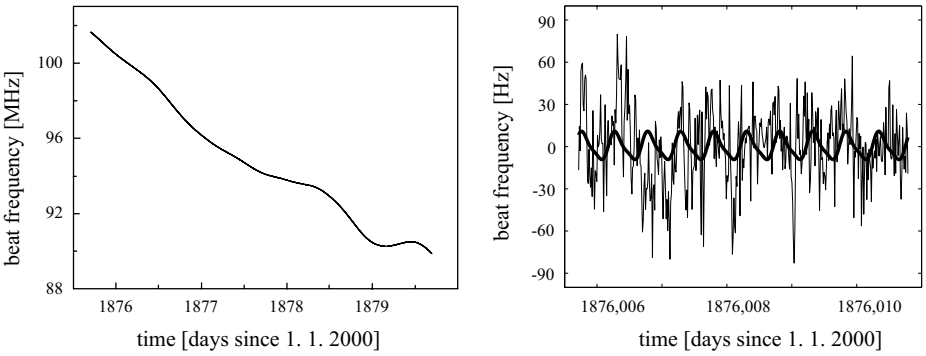


Fig. 6. *Left*: Raw data from a measurement starting February 18. and spanning 4 days. *Right*: A single data subset of this measurement spanning 10 rotations with least squares fit of (15) (fitted drift and offset removed for display)

$$\frac{\Delta\nu}{\nu} = C\cos(2\omega_{\rm rot}T+\phi) + B\sin(2\omega_{\rm rot}T+\phi) + A_0 + A_1 T + A_2 T^2 + A_3\cos(\omega_{\rm rot}T+\phi) + A_4\sin(\omega_{\rm rot}T+\phi)\,. \quad (15)$$

The fixed phase ϕ accounts for an offset in measurement time axis relative to $T = 0$ as defined in Sect. 3. At the chosen subset size including a linear and parabolic drift is sufficient to cleanly distinguish drift from signals at $2\omega_{\rm rot}$. The components at $\omega_{\rm rot}$ are included to yield a proper fit in the presence of residual systematic effects at this frequency. This procedure finally yields a distribution of C and B values for each measurement run containing 200 pairs of values per 24 h.

In a second step we simultaneously fit the time dependence of C and B as modelled by (5) and (6) to these distributions with the Fourier coefficients $C_{\rm ab}$ and $B_{\rm ab}$ as the fit parameters (see Fig. 7). To prevent cross-contamination between different Fourier coefficients we only consider data windows of integer multiples of 24 h for these fits. Discarding excess measurement data is avoided by applying a floating window, subsequently shifted by 1 h, and averaging the results from different windows. Finally from all our measurement runs we extract 15 sets of the 2×5 Fourier coefficients $B_{\rm ab}$ and $C_{\rm ab}$. These are shown in Fig. 8 and the mean values of all 15 results on a certain Fourier coefficient are also stated, weighted according to their fit errors. Note that a constant systematic effect at $2\omega_{\rm rot}$ affects $C_{\rm dc}$ and $B_{\rm dc}$ only. Such an effect is indeed present as apparent from the top most graphs within Fig. 8 and calls for a special consideration of $C_{\rm dc}$ related to κ_{e-}^{ZZ}.

To determine values of SME-parameters involved in these Fourier coefficients the siderial modulation given in Table 1 has to be fitted to these distributions. However at the current span of measurement time of 4 months such a fit does not allow an independent determination of all parameters. To provide at least a first estimate at the current measurement span, we have to either assume the $\tilde{\kappa}_{o+}$- or the $\tilde{\kappa}_{e-}$-elements to be zero and then extract limits on the remaining

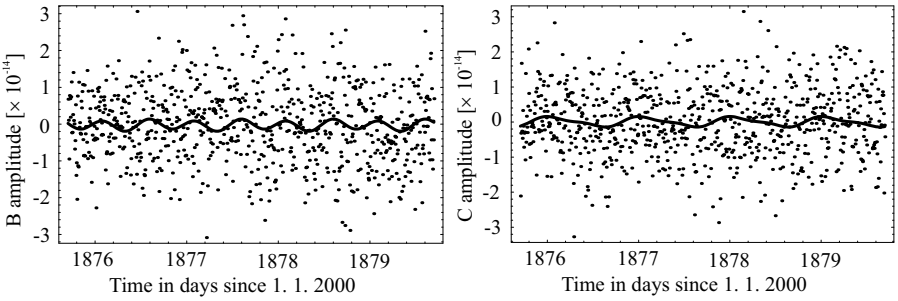


Fig. 7. Resulting distributions for $B(t)$ (*left*) and $C(t)$ obtained from the measurement run shown in Fig. 6. Each point is obtained from a fit of (15) to a subsample spanning 10 rotations. 800 subsamples are included in this measurement run. A least squares fit of (5) respectively (6) is also shown. From these fits the set of 2×5 Fourier coefficients $B_{\rm ab}$ and $C_{\rm ab}$ is obtained

Table 6. SME-parameters extracted from a fit of the relations of Table 1 to the respective distributions of determined Fourier components $B_{\mathrm{ab}}, C_{\mathrm{ab}}$ as shown in Fig. 8. Note that these limits are based on the assumption of no cancellation between $\tilde{\kappa}_{e-}$ and varying $\tilde{\kappa}_{o+}$ terms. All $\tilde{\kappa}_{e-}$ values are $\times 10^{-16}$, $\tilde{\kappa}_{o+}$ values are $\times 10^{-12}$

		ZZ	$XX-YY$	XY	XZ	YZ
$\tilde{\kappa}_{e-}$	this work	−61.8(47.3)	0.4 (5.4)	1.1 (1.8)	4.9 (4.3)	−3.7(3.9)
	from [6]	–	−32 (46)	−57 (23)	−32 (13)	−5 (13)
$\tilde{\kappa}_{o+}$	this work	–	–	−6.5 (4.3)	0.08 (2.3)	0.02 (2.7)
	from [6]	–	–	−18 (15)	−14 (23)	27 (22)

elements from the respective fits. Based on this approximation we obtain the values given in Table 6 reaching down to the low 10^{-16}-level. Compared to the values found by [6] this represents up to an order of magnitude improvement in accuracy.

As stated above the C_{dc} component has to be devoted a special consideration. While this coefficient is compromised by systematic effects, we observe that these average out when considering the distribution of all 15 measurement runs, resulting in an increased error bar only. Taking into account that the $\beta_{\oplus}\kappa_{o+}$-terms can be restricted to a level of 10^{-15} from the remaining coefficients, this allows us to set a limit on the component κ_{e-}^{ZZ} alone of $(-6.2 \pm 4.7) \times 10^{-15}$.

The analysis according to the RMS-framework follows the same steps performed for the SME-analysis. To extract a limit on the RMS-parameter $\mathcal{B}$ from the results for the Fourier coefficients B_{ab} and C_{ab} we perform a weighted least squares fit of the relations of Table 5 to the data shown in Fig. 8, excluding the component C_{dc} compromised by small residual systematic effects. As we neglect siderial modulation of v this is equivalent to extracting $\mathcal{B}$ from a weighted average of the coefficients of Fig. 8. This procedure finally yields $\mathcal{B} = (1.7 \pm 1.8) \times 10^{-10}$, representing an improvement of a factor of eight compared to the result of [3].

6 Outlook

In conclusion, our setup applying precision tilt control proves that comparatively high rotation rate can be achieved at low systematic disturbances. This is a major advance compared to similar past experiments such as the one of [8] and provides the possibility to significantly increase sensitivity of these tests to LLI-violation. From our measurement data currently spanning 4 months we can already set limits on several test theory parameters that are more stringent by up to an order of magnitude compared to previous limits. The main limitation of accuracy within our experimental setup is set by laser lock stability. To improve on this an active vibration isolation as well as new cavities will be implemented within the near future, which should allow us to improve laser lock stability by about an order of magnitude. This should ultimately lead to tests of LLI-violation at a level of 10^{-18}.

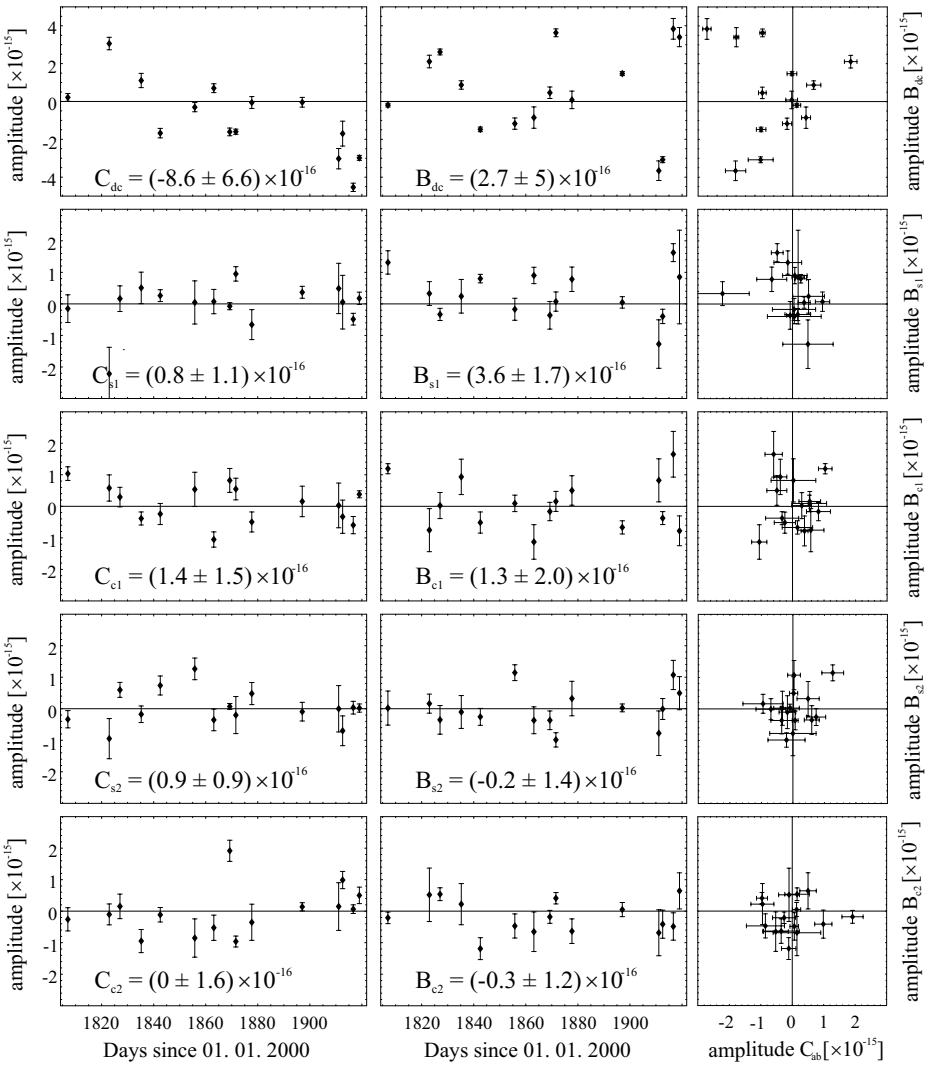


Fig. 8. Each graph gives the results for a certain Fourier coefficient of (5) and (6). Each point within a graph is obtained from one of the 15 measurement runs and the error bar gives the corresponding fit error. The mean values of all 15 results on a certain Fourier coefficient are also stated, weighted according to their fit errors. Note the different scale of the top graphs for C_{dc} and B_{dc} affected by small residual systematic effects

We thank Claus Lämmerzahl for discussions, Jürgen Mlynek and Gerhard Ertl for making this experiment possible. H. Müller acknowledges support from the Alexander von Humboldt–Stiftung. S. Herrmann acknowledges support from the Studienstiftung des deutschen Volkes.

References

1. A.A. Michelson, Am. J. Sci. **22**, 120 (1881); A.A. Michelson and E.W. Morley, Am. J. Sci **34**, 333 (1887); Phil. Mag. **24**, 449 (1897).
2. V.A. Kostelecký and M. Mewes, Phys. Rev. D 66, 056005 (2002).
3. H. Müller et al., Phys. Rev. Lett. **91**, 020401 (2003).
4. H. Müller et al., App. Phys. B **77**, no. 8, pp. 719–731 (2003).
5. P. Wolf et al., Phys. Rev. Lett. **90**, 060402 (2003).
6. P. Wolf et al., Phys. Rev. D **70**, 051902(R) (2004).
7. J.A. Lipa et al., Phys. Rev. Lett. **90**, 060403 (2003).
8. A. Brillet and J.L. Hall, Phys. Rev. Lett. **42**, 549 (1979).
9. P.L. Stanwix et al., Phys. Rev. Lett. **95**, 040404 (2005).
10. P. Antonini et al., Phys. Rev. A **71**, 050101(R) (2005).
11. Comparing the resonance frequencies of two orthogonal rotating cavities increases sensitivity to LLI-violation by a factor of two and thus the original design of the experiment comprised comparison of two rotating cavities. However one of the cavities turned out to be of poor quality after implementation due to a damaged mirror coating. Replacing this resonator by a third resonator was not possible, as the only one available had a length of $L = 10\,\mathrm{cm}$ and could not be fitted into the limited space of the vacuum chamber.
12. R.W.P. Drever et al., Appl. Phys. B **31**, 97–105 (1983).
13. J. Gundlach, priv. comm.; B.R. Heckel, in Proc. of the Second Meeting on CPT and Lorentz Symmetry, Singapore: World Scientific, pp. 173–180 (2002).
14. S.M. Carroll, G.B. Field, R. Jackiw, Phys. Rev. D **41**, 1231 (1990).
15. M.E. Tobar et al., Phys. Rev. D **71**, 025004 (2005).
16. V.A. Kostelecký and M. Mewes, Phys. Rev. Lett. **87**, 251304 (2001).
17. The notation in (5) and (6) has been changed in a straightforward way compared to the notation in [2]
18. H. Müller et al., Phys. Rev. D67, 056006 (2003)
19. H. Müller et al., Phys. Rev. D **68**, 116006 (2003).
20. H. Müller Phys. Rev. D **71**, 045004 (2005).
21. H.P. Robertson, Rev. Mod. Phys. **21**, 378 (1949).
22. R.M. Mansouri and R.U. Sexl, Gen. Rel. Gravit. **8**, 497 (1977); see also C. Lämmerzahl et al., Int. J. Mod. Phys. D **11**, 1109 (2002).
23. To perform a complete test of LLI, i.e. to determine the complete set of test parameters α, β, δ two further experiments are required: The Kennedy–Thorndike experiment [5, 24, 25] is sensitive to boost dependence of $c = c(v)$ and tests a parameter combination $A = (\alpha - \beta + 1)$. The Ives–Stilwell experiment [26, 27] measures the quadratic Doppler effect and determines $(\alpha + \frac{1}{2})$.
24. C. Braxmaier et al., Phys. Rev. Lett. **88**, 010401 (2001)
25. R.J. Kennedy and E.M. Thorndike, Phys. Rev. **42**, 400 (1932).
26. G. Saathoff et al., Phys. Rev. Lett. **91**, 190403 (2003).
27. H.E. Ives and G.R. Stilwell, J. Opt. Soc. Am. **28**, 215 (1938).

A Precision Test of the Isotropy of the Speed of Light Using Rotating Cryogenic Optical Cavities

S. Schiller, P. Antonini, and M. Okhapkin

Institut für Experimentalphysik, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany
step.schiller@uni-duesseldorf.de

Abstract. A test of Lorentz invariance for electromagnetic waves was performed by comparing the resonance frequencies of two stable optical resonators as a function of orientation in space. The crystalline resonators were operated at 3.4 K in a cryostat employing a pulse–tube refrigerator. A new analysis yields the Robertson-Mansouri–Sexl theory parameter combination $\beta - \delta - 1/2 = (-0.6 \pm 2.1 \pm 1.2) \cdot 10^{-10}$ and one parameter of the Standard Model Extension theory, $(\tilde{\kappa}_{e-})^{ZZ} = (-2.9 \pm 2.2) \cdot 10^{-14}$.

1 Introduction

The isotropy of space is a well-tested symmetry of nature [1]. Because it is a foundation of today's accepted theories of the fundamental forces it continues to be the focus of both theoretical and experimental studies. A series of experiments [2–4] have recently been performed with the goal of improving the limits for a hypothetical violation. They were in part motivated by the development of an extension of the Standard Model (SME) by Kostelecký and coworkers [5,6] that describes Lorentz violation in a comprehensive way. This dynamical test theory indicates that isotropy violation, if it exists, may exhibit characteristics that differ from those of previous kinematic test theories, such as the Robertson-Mansouri-Sexl (RMS) theory [7].

Here, we will not discuss the conceptual frameworks used to describe hypothetical violations of isotropy, since this is reported in the literature and is also treated in this Proceedings volume. In this contribution, we limit ourselves to the description of an experiment used to perform an improved test of the isotropy of the speed of light. The experiment has already been presented previously [8]; here we give a more detailed description and report an extension of the data analysis.

The experiment was conceived as an actively rotated Michelson-Morley experiment using ultrastable optical cavities interrogated by lasers. It was a natural extension of our previous work with stationary resonators [3,9], but employed a completely new apparatus, except for the sapphire optical cavities.

S. Schiller et al.: *A Precision Test of the Isotropy of the Speed of Light Using Rotating Cryogenic Optical Cavities*, Lect. Notes Phys. **702**, 401–415 (2006)
DOI 10.1007/3-540-34523-X_14 © Springer-Verlag Berlin Heidelberg 2006

The experiment consists in measuring the difference (beat) $\nu_1 - \nu_2$ between the frequencies of two longitudinal modes of two orthogonal standing-wave cavities as a function of orientation in space. If isotropy is violated, according to the Robertson-Mansouri-Sexl test theory and the Standard Model extension, the beat frequency will vary as

$$\frac{\delta(\nu_1(t) - \nu_2(t))}{\nu} = 2B(t)\sin 2\theta(t) + 2C(t)\cos 2\theta(t) \,, \tag{1}$$

where $\nu_1 \approx \nu_2 \approx \nu$ is the average frequency ($2.8 \cdot 10^{14}$ Hz in this experiment) and $\theta(t)$ is the angle between one cavity's axis relative to the south direction.

In the RMS test theory, the amplitudes $2B(t)$ and $2C(t)$ are proportional to the parameter combination $\beta - \delta - 1/2$, where β, δ parametrize deviations of the frame transformation equations from the usual Lorentz form. The explicit form is given below.

In the SME test theory [5], each amplitude $2B(t)$ and $2C(t)$ is a linear combination of eight coefficients weighted by time-harmonic factors. The amplitude $B(t)$ contains frequency components at 0, $\omega_\oplus$, $2\omega_\oplus$, $\omega_\oplus \pm \Omega_\oplus$ and $2\omega_\oplus \pm \Omega_\oplus$, while $C(t)$ contains in addition one component at the frequency $\Omega_\oplus$. Here $\omega_\oplus$ is Earth's sidereal angular frequency and $\Omega_\oplus$ is Earth's orbital frequency. The determination of the individual $\tilde{\kappa}_{o+}$ coefficients requires the ability to resolve the contribution of Earth's orbital motion in order to discriminate between modulation frequencies differing by $\Omega_\oplus$. Thus a measurement extending over at least 1 year is necessary. However, in this experiment we concentrated on a single parameter, $(\tilde{\kappa}_{e-})^{ZZ}$, and using previous results for the remaining parameters, it was possible to obtain a result within a much shorter measurement time.

2 Experimental Setup

An overall view of the experimental setup is shown in Fig. 1. The whole setup was actively rotated by a computer-controlled precision ball bearing turntable. The turntable itself rested on an optical table (3 m × 1.5 m) that was not floated. An octagonal base plate mounted on the turntable carried most of the components, except for the vacuum forepump, the Helium compressor, and a synthesizer. The cryostat was attached via columns to a plate that could be rested on top of a rack mounted on the base plate. Cryostat and plate could be removed from the rack for opening. On the plate the rotary valve and its driver, temperature controllers, dataloggers, synthesizers and power supplies were installed. The servo systems for the cavity frequency locks and the laser power stabilizations were mounted on the sides of the rack. The laser systems, enclosed by boxes, were located on a breadboard placed on the octagonal base plate. The beat frequency detector was contained in one of the boxes. Thermal insulation (not shown in the figure) was used to shield several of the components of the setup.

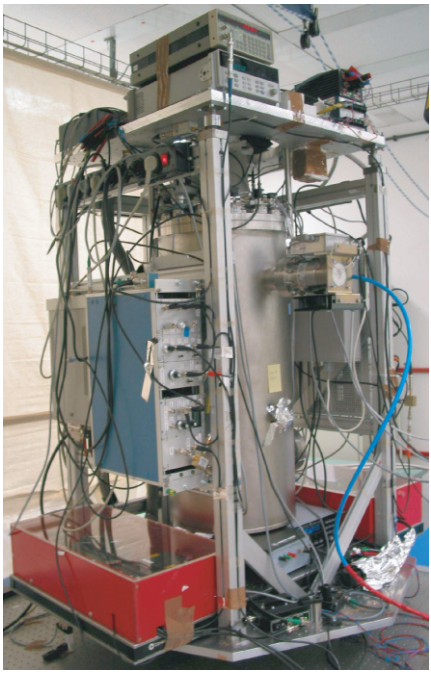


Fig. 1. The experimental setup. The turntable is located under the octagonal base plate and is not visible in the photograph. See text for details

2.1 The Cryogenics

The cryostat is sketched in Fig. 2. Cooling was implemented by a two-stage pulse-tube cooler [10] (Transmit GmbH), using high pressure (18 bar) Helium as working medium. The first stage had a high cooling power (6 W at 50 K) and reached a temperature of approx. 41 K, while the second had a lower cooling power (approx. 0.2 W at 3 K) but was able to reach a minimum temperature of about 2.2 K without load, and 3.2 K when loaded with the experimental set-up used for this experiment. The cooler was driven by a water-cooled Helium compressor (Leybold) with 6 kW power consumption. A significant advantage of this novel cooler type compared to standard cryocoolers is the absence of moving parts inside the cryostat; only the He gas moves within the cryostat, under periodically modulated pressure. Mechanical motion was, however, present in the rotary valve on the top plate. As a consequency, the displacement of the resonators was modulated at the rotary valve frequency, with an amplitude of approx. 1 μm vertically and horizontally, as determined from the propagation of the laser beam exiting the cavities.

An optical cryostat was used, containing three free-space optical access ports: two half-inch diameter windows for horizontal access and an additional window located at the bottom (not used here). Anti-reflection-coated BK7 was used as window material. To avoid backreflections, the windows' normals are angled

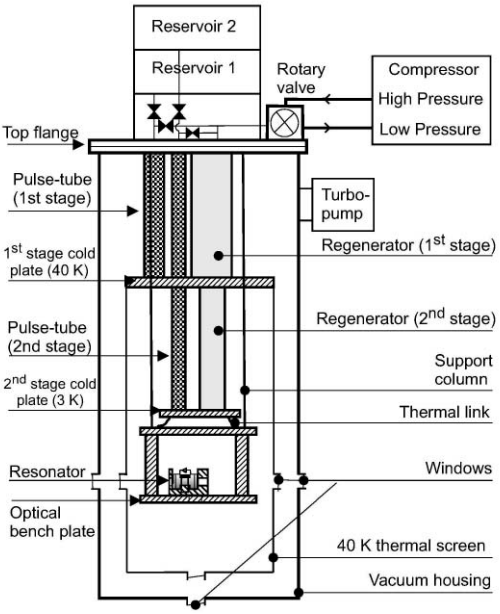


Fig. 2. Cut–away schematic of the cryostat with internal components (not to scale). Height and diameter of the vacuum housing are 90 and 40 cm, respectively. Only one resonator is shown

with respect to the beam direction. The lateral ports were used during the laser alignment and frequency lock phases to identify the cavity modes excited by the lasers.

A copper heat shield (thermal screen) was attached to the first stage cold plate. To improve shielding from the 300 K vacuum can, it also contained angled windows. The space available below the second stage cold plate was about 30 cm in height and 30 cm in diameter. The experimental assembly consisted of an upper plate, four columns, and a bottom optical bench plate, all made of copper, rigidly connected together. It was attached to the top flange by 3 hollow stainless steel rods of 40 cm length, which were heat sunk to the first stage cold plate but not to the second stage cold plate [11]. Copper mesh provided the thermal link between the experimental assembly and the second stage cold plate. The assembly was thermally shielded by superinsulation foils.

A turbo pump was used to continously evacuate the chamber. The top flange of the cryostat contained twelve KF flanges that were used for electrical signal and optical fiber vacuum feedthroughs.

After evacuation of the cryostat, the cool-down time to 3.2 K was about 12 hours.

On a timescale of 10 min, the temperature of the first stage varied by less than 0.1 K. On the optical bench plate, where the cavities are located, the variations were significantly lower, with a 4 mK short-time (10 s) temperature instability and 50 mK instability over long times (10 h). These variations came from (small)

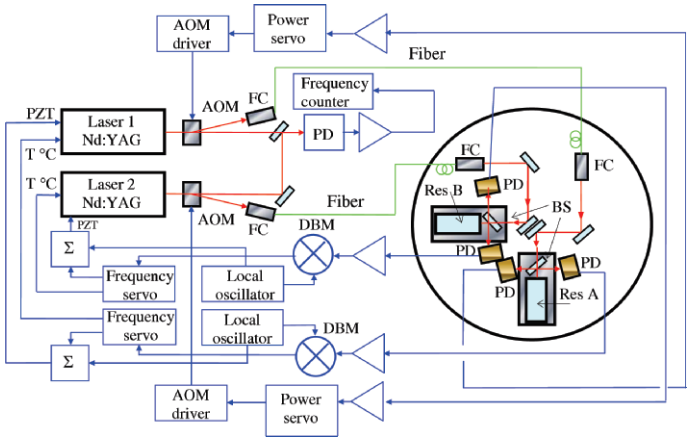


Fig. 3. Schematic of the opto-electronic system. Two Nd:YAG lasers (1064 nm) are frequency-locked to two sapphire optical resonators located in the cryostat. The beams are fed to the resonators via optical fibers. Acousto-optic modulators (AOM), stabilize the power of the beams fed to the resonators. DBM: doubly-balanced mixer; BS: beam splitter; PD: photodiode; FC: fiber coupler, PZT: piezoelectric frequency control actuator; T: temperature control of the laser crystal

instabilities of the pulse-tube cooler itself, and from a dependence of the temperature of the cold stages on the ambient temperature. To reduce these variations, the temperature of the resonators was kept stable at a value slightly above the second stage temperature by active temperature control. A combination of a heater attached to the underside of the optical bench plate, equidistant from the two optical resonators, and a high sensitivity thin film sensor fixed to one resonator housing, together with a commercial digital temperature controller was used, and kept the temperature constant at 3.4 K.

2.2 The Optics Setup

An overview of the functional parts of the laser and resonator system is shown in Fig. 3, and the components on the optics base plate are shown in Fig. 4.

The two optical resonators are made of pure sapphire (Al_2O_3) [12]. Each consists of a 3 cm long cylindrical spacer with inner diameter 1.0 cm, outer diameter 2.6 cm, and crystal c-axis parallel to the cylinder axis. The sapphire mirrors have 1 m radius of curvature and are optically contacted to the spacer. A small hole perpendicular to the cavity axis serves for evacuation. The mirrors are coated for high reflection at 1064 nm. These resonators were already used for relativity tests [3,9,13,14]. The linewidths of the two resonators were 100 kHz at the time of the experiment.

Sapphire was chosen because of its very low thermal expansion coefficient at cryogenic temperature and its low dimensional drift [13]. For dielectric crystals, the thermal expansion is due solely to phonons and therefore the expansion coefficient drops as T^3 as the temperature T approaches zero. For sapphire the

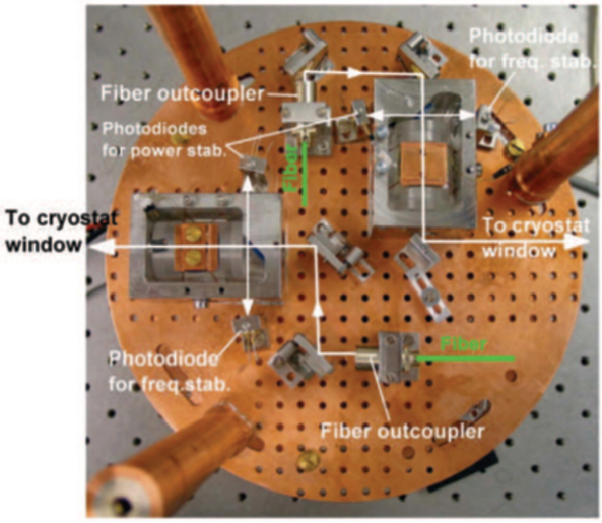


Fig. 4. View of the main cryogenic optical components located on the optics bench plate, without electrical cables and optical fibers. The resonator housings shown were later replaced by gold–plated copper housings

value at 3.4 K is approx. $8 \cdot 10^{-11}\,\mathrm{K}^{-1}$ [15]. We note, however, that the effective expansion coefficient of a mounted cavity can differ substantially, because it also involves the interaction of the resonator with its holder. Thus, in past experiments we have found zero crossings of the thermal expansion coefficient at 3 K [16].

In the present setup, the two resonators were mounted in two housings made of invar and coated with a 5 µm thick gold layer. Gold was used to obtain good contact between the resonators and their housings. The resonators were then fixed to the housing using thin copper straps. The straps were not strongly tightened, to avoid squeezing the resonators because of contraction of the straps during cool-down.

The laser beams were transported to the cavities via two 4 m long polarization-maintaining single-mode fibers with 8° angled fiber ends (to avoid back reflections and miminize etalon effects). The fiber ends were connected to fiber couplers containing short focal length lenses and rigidly attached to the optical base plate. For each laser beam, two adjustable vacuum compatible stainless steel mirror mounts deflected the light toward a 90% reflection/10% transmission beam splitter placed just before the cavity. The light directly reflected from the beam splitter reached a 2 mm diameter InGaAs photodiode that provided the signal for power monitoring and stabilization. The transmitted light was reflected by the cavity, then partially by the beam splitter and was sent to another photodetector of the same type to provide the signal for frequency locking.

The alignment of the laser beams was done at room temperature with the cryostat open. After cooling, the lasers were locked to the TEM_{00} modes, which

could be identified with the help of CCD cameras that monitored the cavity light leaving the cryostat through the windows. The coupling efficiency was typically 10% at room temperature. The efficiency was reduced by a factor 2 in the cold state due to differential thermal contraction effects.

Two diode-pumped monolithic non-planar ring oscillator Nd:YAG lasers emitting 200 mW at 1064 nm were used. The power fed into the fibers was about 1.5 mW. The Drever-Hall reflection locking scheme was used [17]. The lasers were phase modulated at frequencies of 300 kHz via the piezoelectric actuators acting on the laser crystals [18]. The error signals had typical SNR > 10 in a bandwidth of 100 kHz after amplification. They were processed by respective analog servos, each employing a loop for the laser piezoelectric actuator (unity gain coefficient at about 15 kHz) and a slow loop for laser crystal temperature control. The accuracy of the servo electronics was better than 0.1 Hz at 100 s integration time.

The power of each laser beam incident on the cavities was about 50 to 100 μW and was actively stabilized to a relative level of $1 \cdot 10^{-4}$ using an acousto-optical modulator (AOM) placed before the fiber outside the cryostat. The AOMs also served as optical isolators.

On the laser breadboard, two parts of the laser beams were superimposed on a fast photodiode, producing a heterodyne signal at the beat frequency $\nu_1 - \nu_2$, about 700 MHz. This frequency was mixed down with a synthesizer to a frequency of about 10 MHz, to exploit the higher accuracy of the counter at lower frequencies. Both synthesizer and frequency counter were phase-locked to the 5 MHz output of a hydrogen maser.

3 Characterization of the Setup

After cool-down, the beat frequency initially exhibited a drift on the order of 1 Hz/s. After 2 months, this was reduced to 0.02 Hz/s. Our following discussion refers to this stable regime of operation.

A main characteristic of the apparatus, the frequency instability of the beat, is shown in Fig. 5. The root Allan variance (RAV) exhibits a peak at about 7 s. This is due to the modulation of the beat frequency by the pulse-tube cooler with peak-peak amplitude of about 400 Hz. Although the cooler has a mechanical frequency at 1.1 Hz, this frequency was aliased to a lower frequency by the 1 s sampling time of the frequency counter. The minimum RAV ($7 \cdot 10^{-15}$) is attained at $\tau = 20 - 30$ s. At the half-period of the rotation, 300 s, the RAV has increased to $1.4 \cdot 10^{-14}$. The main reasons for this level appear to have been due to a low signal-to-noise ratio of the error signal and to the presence of mechanical/thermal noise from the cooler components and from the laboratory. These RAV values are about an order of magnitude higher than those obtained for the same integration time in the previous, nonrotating, experiment using these resonators [3]. However, for the purpose of the isotropy test the relevant integration time was $\tau = 12$ h in the latter experiment, for which the instability was higher than that at 300 s, relevant in the present experiment.

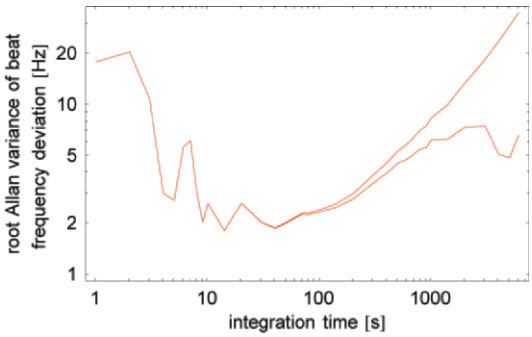


Fig. 5. Root Allan variance of beat frequency under rotation (without tilt correction and temperature decorrelation). *Upper curve*: from raw data including drift; *lower curve*: after removal of the drift (approx. 0.02 Hz/s) from raw data. The time record is the same as in Fig. 7, left

3.1 Laser Power

The dependence of the TEM_{00} resonance frequencies on the power of the beam impinging on the resonators could reliably be measured only for relatively large power changes ($> 5\,\mu W$) and was dependent on the mode-match efficiency. For the resonator with the larger sensitivity, a conservative upper limit is 50 Hz/μW. For the power levels reaching the cavities and the relative power instability given above, this implies an influence of the residual power fluctuations of not more than 0.5 Hz ($1.8 \cdot 10^{-15}$).

3.2 Tilt of the Resonators

The sensitivity of the beat frequency on the orientation of the cryostat was measured.

The experiment was operated on an optical table that was not floated but was supported by a metal frame. The lengths of the feet of the frame could be changed by acting on screws. Before each run, the tilt sensitivities were determined by measuring the frequency shift as function of the inclination of the cryostat in two orthogonal directions. To this end, the leg screws were turned and the orientation of the optical table as a whole was changed. The resulting cryostat inclination was measured by a sensitive two-axis tilt-sensor attached to the plate that holds the cryostat. The resolution of the tilt sensor was 0.1 μrad. The tilt sensor output showed a small dependence on the temperature, but this effect was suppressed by a passive temperature insulation.

We measured a sensitivity of about 0.06 Hz/μrad for tilts around two axes parallel to the resonator axes. Typically, the peak-to-peak tilt variation during rotation was 80 μrad, corresponding to 5 Hz beat modulation. Because of this magnitude, the tilt effects were taken into account in the data analysis.

3.3 Temperature Sensitivities

We measured the dependence of the beat frequency on the temperature of the optical bench by changing the set-point of the temperature controller. The temperature was measured using a Cernox sensor (Lakeshore), connected to the housing of one of the two cavities. Because of the high thermal conductivity of the bench the temperature difference between the two resonators is expected to be very small. The temperature sensitivity was 1.5 Hz/mK. This is equivalent to a thermal expansion coefficient difference between the two cavities of $5.3 \cdot 10^{-12}$/K, a value 15 times lower than the nominal expansion coefficient of sapphire at the same temperature.

The typical residual temperature variation correlated with rotation are 0.15 mK peak-to-peak, leading to an influence on the beat frequency of 0.2 Hz ($0.8 \cdot 10^{-15}$). The instability of the temperature of the optical bench for various time scales is shown in Fig. 6.

Ambient and cryostat component temperatures did have a significant effect on the beat frequency.

The sensitivity of the beat frequency to changes of the temperature in the lab was measured by acting on the air conditioning system. A sensitivity of approx. 75 Hz/K was measured. We did not observe any strong correlation between the lab temperature and the temperature of the optical bench cold plate. When the lab temperature was modulated so as to give 40 Hz beat modulation, the cold plate temperature's peak-peak amplitude was smaller than 0.3 mK, corresponding to a calculated beat frequency modulation of less than 0.5 Hz. Thus, it appears that the observed temperature sensitivity of the beat frequency was due to the thermal sensitivity of the fibers. Temperature changes affected their optical path length and in the presence of residual amplitude modulation and spurious etalons the laser frequency lock point changed. The fiber temperature is influenced both directly by the ambient temperature (detectable by heating

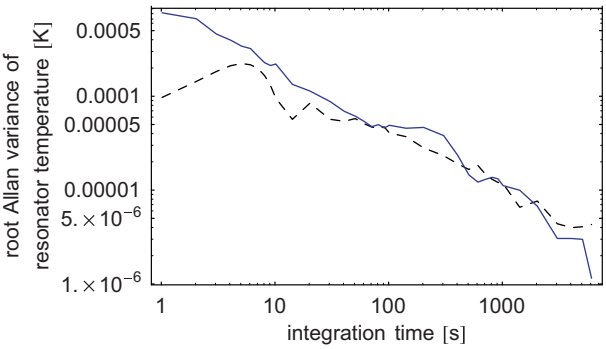


Fig. 6. Root Allan variance of the (actively stabilized) temperature measured near one of the resonators. *Full line*: with rotations: the plateau at 300 s is due to a (residual) modulation correlated with rotation. The time record is the same as in Fig. 7. *Dashed line*: from a record without rotations

the fibers locally) and by their contact to various parts of the cooler inside the cryostat, whose temperatures also changed with ambient temperature and cooler operating conditions.

Typical temperature variations measured at the top of the cryostat during one rotation were as low as 25 mK peak-peak after improvements in the temperature insulation of the setup, corresponding to a 2 Hz peak-to-peak effect assuming the above temperature sensitivity.

In order to characterize to what extent the various temperature variations induced beat frequency variations, we analyzed the correlations. The monitored temperatures were two laboratory temperatures measured at the top and bottom of the cryostat, four temperatures on different cooler components inside the cryostat and the temperature of one resonator holder. A linear regression analysis showed that there are strong correlations between these parameters and the beat frequency. Figure 7 shows an example of this analysis. As can be seen, the slow variations of the beat frequency with respect to a nearly constant drift are to a large extent removed.

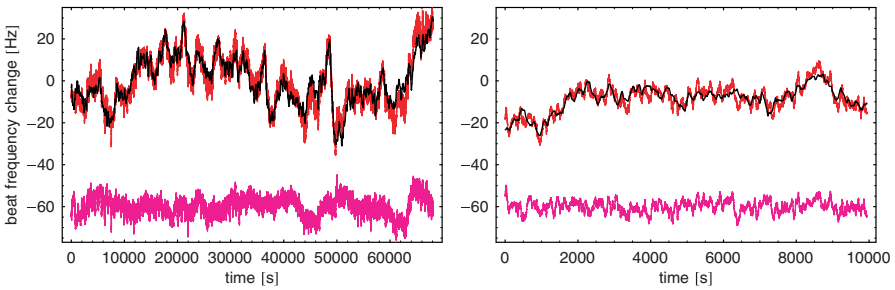


Fig. 7. Influence of temperature variations. *Left*: decorrelation of a 70 ks long beat frequency data set; *right*: decorrelation of a 10 ks subset starting at 50 ks. *Dark gray*: beat frequency, after subtraction of linear and quadratic drift and correction for tilts. *Black*: Linear combination of seven temperature traces that best fits the corrected beat frequency. *Light gray (bottom traces)*: residuals, offset for clarity. During this measurement, the apparatus was rotated. Traces are averaged over 21 s

A summary of the systematic effects is given in Table 1. Not included are sensitivities to the temperatures of individual refrigerator components, since they could not be measured.

4 Data Collection and Analysis

A computer-controlled rotation stage rotated the cryostat over a range of 90°. The total range accessible was limited to a value slightly above this by the He pressure lines connecting the pulse-tube cooler with the stationary compressor.

Table 1. Characterized systematic effects. The column "Inst/mod" lists the measured parameter instability on the timescale of the rotation half-period or the peak-peak modulation. The column "Systematic" is the product of the first column and the sensitivity coefficient. The estimates of the systematics are indicative, since we do not distinguish between angular variations of type $\cos\theta$, $\sin\theta$ and $\cos 2\theta$, $\sin 2\theta$, the two latter types being the relevant ones for the isotropy test

Effect	Inst/Mod	Sensitivity	Systematic	Relative
Tilt	50 µrad	0.06 Hz/µrad	3 Hz	$1.1 \cdot 10^{-14}$
Ambient temperature	0.025 K	75 Hz/K	2 Hz	$0.7 \cdot 10^{-14}$
Resonator temperature	150 µK	1.5 Hz/mK	0.2 Hz	$0.8 \cdot 10^{-15}$
Laser power	10 nW	50 Hz/µW	0.5 Hz	$1.8 \cdot 10^{-15}$

The period of rotation was chosen as 600 s. Shorter periods led to a significant shaking of the cryostat and were therefore not used.

Two synchronized computers collected the data. One computer controlled the rotation angle and rotation speed of the experiment, recorded the beat frequency and the temperature of the optical resonators, by means of a program written in LabView. Sampling time was 1 s. The second computer recorded the cryostat tilt angles, and several temperatures as mentioned above. The tilt and temperature values were used in the data analysis as explained below.

The data discussed here was obtained after about two years of testing and improvements on the whole system. Test runs performed initially typically exhibited significant drifts of properties of the apparatus, such as cooler internal temperatures. After minimizing these variations, we succeeded in obtaining stable operation.

For data analysis, the (aliased) beat frequency modulation was removed from the data by filtering in the Fourier domain, and the sum of the tilt angles multiplied by the respective tilt sensitivities was subtracted. Decorrelation of the temperatures was then performed, if desired.

The beat frequency in each interval $\theta = [0°; 90°; 0°]$ (labeled by i) was least-squares fitted with the three-parameter function

$$a_i t + 2B(t_i)\sin 2\theta(t) + 2C(t_i)\cos 2\theta(t) \ , \tag{2}$$

where the coefficient a_i quantifies a linear drift that may vary from rotation to rotation. An example of an analysis of a single rotation is shown in Fig. 8. The obtained amplitude sets $\{2B(t_i)\}$, $\{2C(t_i)\}$ are then analyzed according to the Robertson-Mansouri-Sexl test theory and the Standard Model Extension.

4.1 Analysis in the RMS Framework

It can be shown [8] that according to the RMS test theory, the amplitudes of the beat frequency modulation with angle are given by

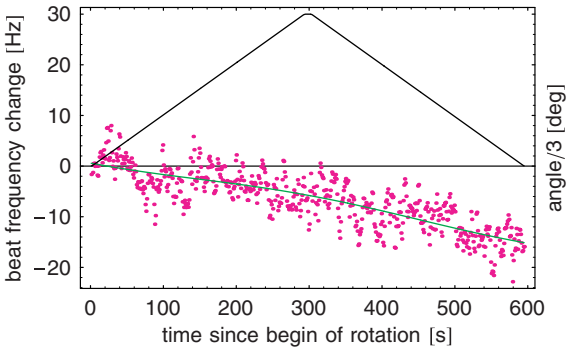


Fig. 8. Beat frequency as a function of rotation of the apparatus. *Gray dots*: tilt-corrected beat frequency, cooler-induced modulation removed, no averaging, 1 s sampling time; *gray line*: fit of beat frequency according to (2) with values $2C\nu = -0.76 \pm 0.18\,\mathrm{Hz}, 2B\nu = -0.02 \pm 0.41\,\mathrm{Hz}$; *black line*: angular position

$$2B(t) = (1/2 - \beta + \delta)(v^2/c_0^2)(\gamma_3 \cos\omega_\oplus T_\oplus + \gamma_4 \cos 2\omega_\oplus T_\oplus + \sigma_3 \sin\omega_\oplus T_\oplus + \sigma_4 \sin 2\omega_\oplus T_\oplus)\ , \tag{3}$$

$$2C(t) = (1/2 - \beta + \delta)(v^2/c_0^2)(\gamma_0 + \gamma_1 \cos\omega_\oplus T_\oplus + \gamma_2 \cos 2\omega_\oplus T_\oplus + \sigma_1 \sin\omega_\oplus T_\oplus + \sigma_2 \sin 2\omega_\oplus T_\oplus)\ , \tag{4}$$

where the constants are defined in Table 2.

$T_\oplus$ is the time since the beginning of the data plus an offset that accounts for a time difference since the coincidence of the lab's y axis with the $\hat{Y}$ axis of the Sun-centered system [5]. The y axis is parallel to one cavity axis when in the 0° position. The direction of the Sun's velocity $\mathbf{v}$ relative to the cosmic microwave background is given by the right ascension $\Phi = 168°$ and the declination $\Theta = -6°$.

In order to provide an analysis similar to that of [19, 20], we consider a measurement interval extending over 183 hours that contained 940 rotations (after removal of a small number of outliers), grouped in 5 sets. Only the tilt correction was implemented and each set was fitted with the functions (3) and (4) plus additional contributions $b_{syst} + b'_{syst}t_i$ and $c_{syst} + c'_{syst}t_i$ that model

Table 2. The values of γ_i and σ_i appearing in (3) and (4)

$\gamma_0 = \frac{1}{4}\sin^2\chi(3\cos^2\Theta - 1)$	
$\gamma_1 = \frac{1}{2}\cos\Phi\sin 2\Theta\sin 2\chi$	$\sigma_1 = \gamma_1\tan\Phi$
$\gamma_2 = \frac{1}{4}\cos 2\Phi\cos^2\Theta(\cos 2\chi - 3)$	$\sigma_2 = \gamma_2\tan 2\Phi$
$\gamma_3 = \sigma_3\tan\Phi$	$\sigma_3 = \cos\Phi\sin\chi\sin 2\Theta$
$\gamma_4 = -\sigma_4\tan 2\Phi$	$\sigma_4 = \cos^2\Theta\cos\chi\cos 2\Phi$

systematic effects. The individual sets yield the following values and standard errors for $\beta - \delta - 1/2$: $\{(5.5, 5.0), (7.2, 9.0), (-8.3, 3.7), (5.7, 4.5), (-2.6, 3.9)\} \cdot 10^{-10}$. The overall result is

$$\beta - \delta - 1/2 = (-0.6 \pm 2.1 \pm 1.2) \cdot 10^{-10} \quad , \tag{5}$$

where the first error is statistical and the second error reflects the uncertainty in the experimentally determined tilt sensitivities and an estimate of the influence of laser power variations.

4.2 SME Test Theory

Considering that the time span over which data was taken was significantly less than one year, the main goal of the data analysis within the SME model was the determination of a value for $(\tilde{\kappa}_{e-})^{ZZ}$. The results of the cryogenic microwave experiment [4] found the elements of $(\tilde{\kappa}_{e-})$ (except for $(\tilde{\kappa}_{e-})^{ZZ}$) and the elements of $\beta_{\oplus}(\tilde{\kappa}_{o+})$ to be at most several parts in 10^{-15} in magnitude. If we assume these elements to be zero, we can use the result for $2C$ only to determine $(\tilde{\kappa}_{e-})^{ZZ}$,

$$(\tilde{\kappa}_{e-})^{ZZ} = \frac{4\langle 2C \rangle}{3 \sin^2 \chi} \ . \tag{6}$$

For the set of rotations under the most stable conditions (of 76 hours duration, also considered in [8]) the average is $\langle 2C\nu \rangle = -2.4\,\mathrm{Hz}$ with a sample standard deviation of 1.9 Hz, and $\langle 2B\nu \rangle = 0.8\,\mathrm{Hz}$, with a sample standard deviation 2.6 Hz. In this analysis, the frequency data was also decorrelated with respect to the temperature data, whereby data intervals of 10 ks were decorrelated individually, in order to allow for changing environmental conditions. Without decorrelation of the temperatures the values are, in the same order, (−3.3 Hz, 2.3 Hz) and (2.8 Hz, 2.4 Hz) [8].

We used the large number of rotations performed under different experimental conditions to estimate the uncertainty in $2C$ due to (identifiable) systematic effects at 1.8 Hz. This includes the errors due to the uncertainty of the tilt coefficients and due to laser power variations. This results in[1]

$$(\tilde{\kappa}_{e-})^{ZZ} = (-2.9 \pm 2.2) \cdot 10^{-14} \quad , \tag{7}$$

where the uncertainty is dominated by the systematic effects.

5 Conclusions

The experiment described in this work was performed in order to improve the previous rotating laser experiment by Brillet and Hall, exploiting some of advances in laser stabilization techniques developed since. Our experiment yielded a

[1] In [8], a factor 2 was inadvertently omitted when calculating $(\tilde{\kappa}_{e-})^{ZZ}$ from $2C$.

significant improvement. Similar to their experiment, a strong limit for $\beta-\delta-1/2$ required exploiting the modulation by Earth's rotation. At the same time the present experiment provided an approach to the task of measuring $(\tilde{\kappa}_{e-})^{ZZ}$ with high accuracy (a weaker limit on this quantity can also be extracted from the experiment of Brillet and Hall).

Limitations of the experiment were the sensitivity of the optical path length to temperature, and the limited laser lock quality as a consequence of the relatively weak cavity throughput. This made the beat frequency more sensitive to optical path length variations. In addition, a certain level of mechanical noise was present. In an upgraded experiment, obvious improvements are resonators of higher finesse and throughput, optical path length stabilization, and rotations with shorter period and lower tilt modulation.

In discussing a "null" experiment, it may be argued that a nonzero value of the measured parameter may have been (partially) canceled by an unknown systematic effect, so that the bounds provided by an experiment may be questioned. We emphasize the importance of the fact that three experiments [8, 19, 20] have recently been reported whose results are consistent with each other. Because they were performed by independent groups with different techniques, it is unlikely that they all exhibit a strong cancellation between the respective systematics and a substantial nonzero value of $(\tilde{\kappa}_{e-})^{ZZ}$.

Acknowledgments

We thank E. Göklü for his participation in this work, L. Haiberger for contributions to the cryostat development, A. Nevsky and C. Lämmerzahl for discussions, and G. Thummes for his helpful assistance. P.A. was supported by a DAAD fellowship, M.O. by a Heinrich–Hertz Foundation fellowship. This research was part of the Gerhard–Hess Program of the German Science Foundation.

References

1. A. Brillet and J.L. Hall. Improved laser test of the isotropy of space. *Phys. Rev. Lett.*, 42:549, 1979.
2. J.A. Lipa, J.A. Nissen, S. Wang, D.A. Stricker, and D. Avaloff. New limit on signals of Lorentz violation in electrodynamics. *Phys. Rev. Lett.*, 90:060403, 2003.
3. H. Müller, S. Herrmann, C. Braxmaier, S. Schiller, and A. Peters. Modern Michelson-Morley experiment using cryogenic optical resonators. *Phys. Rev. Lett.*, 91:020401, 2003.
4. P. Wolf, S. Bize, A. Clairon, G. Santarelli, M.E. Tobar, and A.N. Luiten. Improved test of Lorentz invariance in electrodynamics. *Phys. Rev. D*, 70:051902(R), 2004.
5. V.A. Kostelecký and M. Mewes. Signals for Lorentz violation in electrodynamics. *Phys. Rev. D*, 66:056005, 2002.
6. V.A. Kostelecký and M. Mewes. Cosmological constraints on Lorentz violation in electrodynamics. *Phys. Rev. Lett.*, 87:251304, 2001.

7. R. Mansouri and R. Sexl. A test theory of special relativity: I. Simultaneity and clock synchronisation. *Gen. Rel. and Grav.*, 8:497, 1977.
8. P. Antonini, M. Okhapkin, E. Göklü, and S. Schiller. Test of constancy of speed of light with rotating cryogenic optical resonators. *Phys. Rev. A*, 71:050101(R), 2005.
9. C. Braxmaier, H. Müller, O. Pradl, J. Mlynek, A. Peters, and S. Schiller. Test of relativity using a cryogenic optical resonator. *Phys. Rev. Lett.*, 88:010401, 2002.
10. C. Wang and et al. A two-stage pulse tube cooler operating below 4 k. *Cryogenics*, 37:159, 1997.
11. C. Lienerth, G. Thummes, and C. Heiden. Progress in low noise cooling performance of a pulse-tube cooler for HT–SQUID operation. *IEEE Trans. on Applied Superconductivity*, 11:812, 2001.
12. S. Seel, R. Storz, G. Ruoso, J. Mlynek, and S. Schiller. Cryogenic Optical Resonators: A new tool for laser frequency stabilization at the 1 hz level. *Phys. Rev. Lett.*, 78:4741, 1997.
13. R. Storz, C. Braxmaier, K. Jäck, O. Pradl, and S. Schiller. Ultrahigh long-term dimensional stability of a sapphire cryogenic optical resonator. *Opt. Lett.*, 23:1031, 1998.
14. H. Müller, C. Braxmaier, S. Hermann, O. Pradl, C. Lämmerzahl, J. Mlynek, S. Schiller, and A. Peters. Testing the foundation of relativity using cryogenic optical resonators. *IJMPD*, 11:1101, 2002.
15. M. Lucht, M. Lerche, H.C. Wille, Y.V. Shvyd'ko, H.D. Rüter, E. Gerdau, and P. Becker. Precise measurement of the lattice parameters of α-Al_2O_3 in the temperature range 4.5-250 K using the Mössbauer wavelength standard. *J. Appl. Cryst.*, 36:1075, 2003.
16. R. Storz. PhD thesis, Universität Konstanz, Germany, 1998.
17. R.W.P. Drever, J.L. Hall, F.V. Kowalski, J. Hough, G.M. Ford, A.J. Munley, and H. Ward. Laser phase and frequency stabilization using an optical resonator. *Appl. Phys. B*, 31:97, 1983.
18. G. Cantatore, F. Della Valle, E. Milotti, P. Pace, E. Zavattini, E. Polacco, F. Perrone, C. Rizzo, G. Zavattini, and G. Ruoso. Frequency locking of a Nd:YAG laser using the laser itself as the optical phase modulator. *Rev. Sci. Instrum.*, 66:2785, 1994.
19. P.L. Stanwix, M.E. Tobar, P. Wolf, M. Susli, C.R. Locke, E.N. Ivanov, J. Winterflood, and F. van Kann. Test of Lorentz invariance in electrodynamics using rotating cryogenic sapphire microwave oscillators. *Phys. Rev. Lett.*, 95:040404, 2005.
20. S. Herrmann, A. Senger, E. Kovalchuk, H. Müller, and A. Peters. Test of the isotropy of the speed of light using a continuously rotating optical resonator. *arXiv*, physics:0508097, 2005.

Rotating Resonator-Oscillator Experiments to Test Lorentz Invariance in Electrodynamics

M.E. Tobar[1], P.L. Stanwix[1], M. Susli[1], P. Wolf[2,3], C.R. Locke[1] and E.N. Ivanov[1]

[1] University of Western Australia, School of Physics M013, 35 Stirling Hwy., Crawley 6009 WA, Australia
mike@physics.uwa.edu.au
[2] SYRTE, Observatoire de Paris, 61 Av. de l'Observatoire, 75014 Paris, France
[3] Bureau International des Poids et Mesures, Pavillon de Breteuil, 92312 Sèvres Cedex, France

1 Introduction

The Einstein Equivalence Principle (EEP) is a founding principle of relativity [1]. One of the constituent elements of EEP is Local Lorentz Invariance (LLI), which postulates that the outcome of a local experiment is independent of the velocity and orientation of the apparatus. The central importance of this postulate has motivated tremendous work to experimentally test LLI. Also, a number of unification theories suggest a violation of LLI at some level. However, to test for violations it is necessary to have an alternative theory to allow interpretation of experiments [1], and many have been developed [2–7]. The kinematical frameworks (RMS) [2,3] postulate a simple parameterization of the Lorentz transformations with experiments setting limits on the deviation of those parameters from their values in special relativity (SR). Because of their simplicity they have been widely used to interpret many experiments [8–11]. More recently, a general Lorentz violating extension of the standard model of particle physics (SME) has been developed [6] whose Lagrangian includes all parameterized Lorentz violating terms that can be formed from known fields.

This work analyses rotating laboratory Lorentz invariance experiments that compare precisely the resonant frequencies of two high-Q factor (or high finesse) cavity resonators. High stability electromagnetic oscillatory fields are generated by implementing state of the art frequency stabilization systems with the narrow line width of the resonators. Previous non-rotating experiments [10,12,13] relied on the rotation of the Earth to modulate putative Lorentz violating effects. This is not optimal for two reasons. Firstly, the sensitivity to Lorentz violations is proportional to the noise of the oscillators at the modulation frequency, typically best for periods between 10 and 100 seconds. Secondly, the sensitivity is proportional to the square root of the number of periods of the modulation signal, therefore taking a relatively long time to acquire sufficient data. Thus, by

M.E. Tobar et al.: *Rotating Resonator-Oscillator Experiments to Test Lorentz Invariance in Electrodynamics*, Lect. Notes Phys. **702**, 416–450 (2006)
DOI 10.1007/3-540-34523-X_15
© Springer-Verlag Berlin Heidelberg 2006

rotating the experiment the data integration rate is increased and the relevant signals are translated to the optimal operating regime [14].

In this work we outline the two most commonly used test theories (RMS and SME) for testing LLI of the photon. Then we develop the general framework of applying these test theories to resonator experiments with an emphasis on rotating experiments in the laboratory. We compare the inherent sensitivity factors of common experiments and propose some new configurations. Finally we apply the test theories to the rotating cryogenic experiment at the University of Western Australia, which recently set new limits in both the RMS and SME frameworks [15]. Note added: Two other concurrent experiments have set some similar limits [16, 17].

2 Common Test Theories to Characterize Lorentz Invariance

The most famous test of LLI (or the constancy of the speed of light) was that conducted by Michelson and Morley in 1887 [18] with a rotating table and a Michelson interferometer. In actual fact, the theoretical framework used by Michelson and Morley was not a test of LLI, since the concept did not exist at the time, but that of an aether drift. The relative motion of the apparatus through the aether was thought to induce a phase difference between the arms of the interferometer (and hence an interference pattern) depending on the orientation. Thus, as the Earth moved from one end of its orbit to the opposite end, the change in its velocity should be a detectable value. Michelson and Morley found no fringe shifts due to Earth motion around the sun and reported a null result. Since the Michelson Morley experiment, there have been many other types of experiments devised to test the validity of SR and the constancy of light. However, to interpret these experiments one must formulate an alternative test theory, and in this section we outline two of the most commonly used.

2.1 Robertson, Mansouri, Sexl Framework

A simple kinematic test theory that has been widely used is that of Robertson, Mansouri and Sexl (RMS) [2, 3], where time standards ("clocks") and length standards ("rods") are considered without taking into account their underlying structure. This framework postulates a preferred frame $\Sigma(T, \mathbf{X})$ which satisfies LLI, and a moving frame $S(t, \mathbf{x})$, which does not. The prime candidate for the preferred frame is taken as the Cosmic Microwave Background (CMB), since any anticipated non-symmetries are expected to arise from Planck-scale effects during the creation of the universe. In this framework we analyse the Poynting vector direction of the electromagnetic signal with respect to the velocity of the lab through the CMB.

The normal Lorentz Transformations for a boost in the x direction are expressed in a special form below (where c is the speed of light in the Σ frame):

$$dT = \frac{1}{a}\left(dt + \frac{vdx}{c^2}\right); dX = \frac{dx}{b} + \frac{v}{a}\left(dt + \frac{vdx}{c^2}\right); dY = \frac{dy}{d}; dZ = \frac{dz}{d}; \quad (1)$$

Here we take a Taylor expansion for a, b and d of the form: $a \approx 1 + \alpha v^2/c^2 + \mathcal{O}(c^{-4})$; $b \approx 1 + \beta v^2/c^2 + \mathcal{O}(c^{-4})$; $d \approx 1 + \delta v^2/c^2 + \mathcal{O}(c^{-4})$. Recalling that $\gamma = \frac{1}{\sqrt{1-v^2/c^2}}$ from Special Relativity (SR), we see that SR predicts $\alpha = -1/2$ and $\beta = 1/2$. Since SR predicts no contraction in directions orthogonal to a boost, it also predicts that $\delta = 0$. Thus, the RMS parameterizes a possible Lorentz violation by a deviation of the parameters (α, β, δ) from the SR values $(-\frac{1}{2}, \frac{1}{2}, 0)$.

By manipulating (1) to form the infinitesimals in the S frame, we can separate the equation into a boost term $(\beta - \alpha - 1)$, anisotropy term $(\delta - \beta + \frac{1}{2})$ and time dilation parameter $\alpha + \frac{1}{2}$. Thus, a complete verification of LLI in the RMS framework [2,3] requires a test of (i) the isotropy of the speed of light (measuring $P_{MM} = \delta - \beta + \frac{1}{2}$), a Michelson-Morley (MM) experiment [18], (ii) the boost dependence of the speed of light (measuring $P_{KT} = \beta - \alpha - 1$), a Kennedy-Thorndike (KT) experiment [19] and (iii) the time dilation parameter (measuring $P_{IS} = \alpha + \frac{1}{2}$), an Ives-Stillwell (IS) experiment [20, 21]. Rotating experiments may be considered Michelson-Morley experiments and only measure P_{MM}, so in this section we restrict ourselves to these types of measurements.

Assuming only a MM type Lorentz violation, and setting $ds^2 = c^2dT^2 - dX^2 - dY^2 - dZ^2 = 0$ in Σ, and transforming according to (1) we find the coordinate travel time of a light signal in S becomes;

$$dt = \frac{dl}{c}\left(P_{MM} \times \sin^2\theta \frac{v^2}{c^2}\right) + \mathcal{O}(c^{-4}) \quad (2)$$

where $dl = \sqrt{dx^2 + dy^2 + dz^2}$ and θ is the angle between the Poynting vector and the velocity $\mathbf{v}$ of S in Σ. For a modern MM experiment that measures the difference frequency between two resonant cavities, the fractional frequency difference may be calculated from (2) in a similar way to [11] to give:

$$\frac{\Delta\nu_0}{\nu_0} = \frac{P_{MM}}{2\pi c^2}\left[\oint \left(\mathbf{v}.\hat{\mathbf{I}}_{\mathbf{a}}(q_a)\right)^2 dq_a - \oint \left(\mathbf{v}.\hat{\mathbf{I}}_{\mathbf{b}}(q_b)\right)^2 dq_b\right] \quad (3)$$

Where $\hat{\mathbf{I}}_j(q_j)$ is the unit vector in the direction of light propagation (Poynting vector) of each resonator (labeled by subscripts a and b), and q_j is the variable of integration around the closed path coordinates of the Poynting vector of each resonator.

To calculate the relevant time dependent expressions for $\mathbf{v}$, velocities are transformed to a geocentric non-rotating (with respect to distant stars) reference frame (denoted as the MM-frame) centered at the center of mass of the Earth with its z-axis perpendicular to the equator, pointing north, the x-axis pointing toward 11.2h right ascension (aligned with the equatorial projection of $\mathbf{u}$ defined below). A pictorial representation of the frame is shown in Fig. 1. Classical (Galilean) transformations for the velocities and $\sin^2\theta$ are sufficient as relativistic

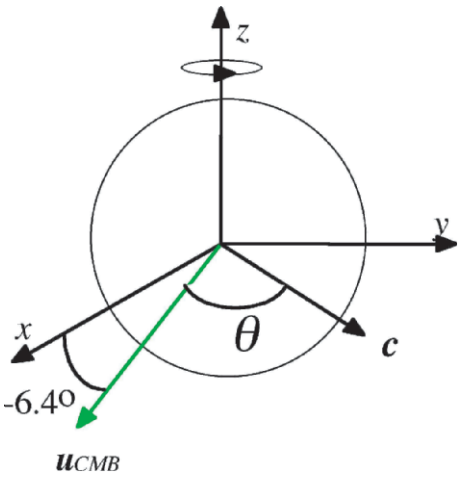


Fig. 1. This frame is an Earth-centered frame in which the spin axis of the earth is the z-axis, and the velocity of the Sun with respect to the CMB is defined to have no component in the y direction. Thus, the Earth is spinning at the sidereal rate within this frame, and the angle θ is shown pictorially, but in general is a function of position and time as the Earth spins and the experiment rotates

terms are of order $\mathcal{O}(c^{-2})$ and therefore give rise in (2) to terms of order $\mathcal{O}(c^{-5})$. We consider two velocities, the velocity of the sun with respect to the CMB $\mathbf{u}$ (declination $-6.4°$, right ascension 11.2 h) and the orbital velocity of the Earth $\mathbf{v}_o$. Velocities due to the spinning of the Earth and laboratory are much smaller and do not impact on the calculations and may be ignored. Thus, the sum of the two provide the velocity of the laboratory in the universal frame to be inserted in (3). In the MM-Earth frame, the CMB velocity is:

$$\mathbf{u} = u \begin{pmatrix} \cos\phi_\mu \\ 0 \\ \sin\phi_\mu \end{pmatrix} \tag{4}$$

where $u \approx 377\,\mathrm{km/s}$ and $\phi_\mu \approx -6.4°$. To calculate the orbital velocity we first consider the Earth in a barycentric non-rotating frame (BRS) with the z-axis perpendicular to the Earth's orbital plane and the x-axis pointing toward 0° right ascension (pointing from the Sun to the Earth at the moment of the autumn equinox).

$$\mathbf{v}_o^{BRS} = v_o \begin{pmatrix} -\sin\lambda_0 \\ \cos\lambda_0 \\ 0 \end{pmatrix} \tag{5}$$

where $v_o \approx 29.78\,\mathrm{km/s}$ is the orbital speed of the Earth, and $\lambda_0 = \Omega_\oplus(t - t_o)$ with $\Omega_\oplus \approx 2.0\ 10^{-7}\mathrm{rad/s}$ the angular orbital velocity and $t - t_0$ the time since the autumnal equinox. We first transform to a geocentric frame (GRS) that has

its x-axis aligned with the BRS one but its z-axis perpendicular to the equatorial plane of the Earth

$$\mathbf{v}_o^{GRS} = \begin{pmatrix} v_o^{xBRS} \\ v_o^{yBRS}\cos\varepsilon - v_o^{zBRS}\sin\varepsilon \\ v_o^{yBRS}\sin\varepsilon + v_o^{zBRS}\cos\varepsilon \end{pmatrix} \tag{6}$$

where $\varepsilon \approx 23.27^\circ$ is the angle between the equatorial and orbital planes of the Earth. We then transform to the MM-Earth frame:

$$\mathbf{v}_o = \begin{pmatrix} v_o^{xGRS}\cos\alpha_\mu + v_o^{yGRS}\sin\alpha_\mu \\ -v_o^{xGRS}\sin\alpha_\mu + v_o^{yGRS}\cos\alpha_\mu \\ v_o^{zGRS} \end{pmatrix} \tag{7}$$

where $\alpha_\mu \approx 167.9^\circ$ is the right ascension of $\mathbf{u}$. Summing the two velocities from (4) and (7) we obtain the velocity of the lab with respect to the "universe rest frame", transformed to the MM-Earth frame

$$\mathbf{v} = \begin{pmatrix} u\cos\phi_\mu + v_o(-\sin\lambda_0\cos\alpha_\mu + \cos\lambda_0\sin\alpha_\mu\cos\varepsilon) \\ v_o(\sin\lambda_0\sin\alpha_\mu + \cos\lambda_0\cos\alpha_\mu\cos\varepsilon) \\ u\sin\phi_\mu + v_o\cos\lambda_0\sin\varepsilon \end{pmatrix}. \tag{8}$$

Substituting in the numeric values gives an orbital velocity of (in m/s);

$$\mathbf{v} = \begin{pmatrix} 374651 + 5735\cos(\lambda_0) + 29118\sin(\lambda_0) \\ -26750\cos(\lambda_0) + 6242\sin(\lambda_0) \\ -42024 + 11765\cos(\lambda_0) \end{pmatrix}. \tag{9}$$

The last calculation to make is the time dependence of the the unit vector $\hat{\mathbf{l}}$ along the direction of light propagation, which will depend on the configuration of the experiment, including the type of resonator and whether it is rotating in the laboratory or not. In Sect. 5 we calculate this dependence for a specific experiment, which uses WG modes rotating in the laboratory.

2.2 Standard Model Extension

The Standard Model Extension (SME) [6] conglomerates all possible Lorentz-Violating terms and incorporates them in a framework, which is an extension of the Standard Model of Particle Physics. There are numerous Lorentz-violating terms per particle sector (i.e. fermions, bosons and photons). However in this work we are restricted to the so called minimal "photon-sector", which only includes 19 terms. The SME adds additional terms to the Lagrangian of the Standard Model for photons. Where as the standard Lagrangian was simply:

$$\mathcal{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} \tag{10}$$

Under the SME, it becomes [7]:

$$\mathcal{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} - \frac{1}{4}(k_F)_{\kappa\lambda\mu\nu}F^{\kappa\lambda}F^{\mu\nu} + \frac{1}{2}(k_{AF})^{\kappa}\epsilon_{\kappa\lambda\mu\nu}A^{\lambda}F^{\mu\nu} \tag{11}$$

where A^{λ} is the 4-potential. The $(k_{AF})^{\kappa}$ terms have the dimensions of mass, and are the CPT odd terms [22]. It is argued in [6] that these should be zero because they induce instabilities as they are non-negative in the Lagrangian. There are also astronomical measurements [7] which place stringent limits on k_{AF}. From here on these terms are set to zero.

On the other hand, the $(k_F)_{\kappa\lambda\mu\nu}$ terms are CPT even, dimensionless and have 19 independent terms out of the 256 possible combinations of κ, λ, μ and ν. Out of these independent Lorentz violating terms, 10 combinations have been analysed using astrophysical polarisation tests and have an upper-limit of 2×10^{-32} [7]. This limit is many orders of magnitude less than what is expected from laboratory experiments, so these terms are set to zero to simplify the calculations and to remain consistent with previous results.

We can derive the equations of motion for this system by minimising the action given by (11), using variational techniques and the definition $F^{\mu\nu} \equiv \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu}$ and $A_{\mu} \equiv (\phi, \mathbf{A})$. These equations are similar to those of a Maxwellian model in anisotropic media instead of a vacuum. In order to express these in a convenient form, we form linear combinations of the CPT even term. These are given below [7]:

$$\begin{aligned} &\kappa_{DE}{}^{jk} = -2(k_F)^{0j0k}; \kappa_{HB}{}^{jk} = \frac{1}{2}\epsilon^{jpq}\epsilon^{krs}(k_F)^{pqrs}; \\ &\kappa_{DB}{}^{jk} = -\kappa_{HE}{}^{kj} = (k_F)^{0jpq}\epsilon^{kpq}. \end{aligned} \tag{12}$$

The dynamics of the model can be described in terms of equivalent **B**, **E**, **H** and **D** fields [7,11] in a vacuum using the matrices in (12):

$$\begin{pmatrix} \mathbf{D} \\ \mathbf{H} \end{pmatrix} = \begin{pmatrix} \epsilon_0(1+\kappa_{DE}) & \sqrt{\frac{\epsilon_0}{\mu_0}}\kappa_{DB} \\ \sqrt{\frac{\epsilon_0}{\mu_0}}\kappa_{HE} & \mu_0^{-1}(1+\kappa_{HB}) \end{pmatrix} \begin{pmatrix} \mathbf{E} \\ \mathbf{B} \end{pmatrix} \tag{13}$$

Note that (13) is rank 6, as the κ matrices are rank 3 as defined in (12). The standard Maxwell equations in a vacuum are recovered if these κ matrices are set to zero.

Thus the effect of the SME in the photon-sector can be interpreted as introducing medium-like properties to the vacuum. In the full SME, this is considered as an effect from Planck-scale physics in the early universe. The κ matrices are all position dependent and thus act as "values" positioned throughout space. If one or more of these values are zero, it does not imply the rest are also zero as there is no relation between each of the independent components. However, there is a linear combination of these components which allows us to separate them into birefringent [7] and non-birefringent terms. By eliminating those values which have been constrained beyond what we hope to achieve in this experiment, these terms can be simply written as in (14).

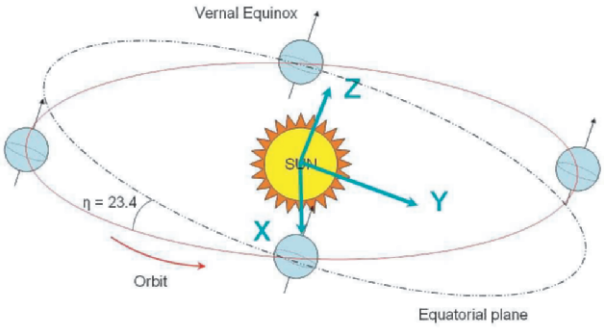


Fig. 2. The Sun-Centered Celestial Equatorial Frame (SCCEF), with the Earth at different equinoxes during the year. Note that during the vernal equinox, the longitude which is at noon has it's Y axis pointing east, while the longitude which is at midnight has it's Y axis pointing west, and vice-versa for the autumn equinox

$$
\begin{aligned}
(\widetilde{\kappa}_{e+})^{jk} &= \frac{1}{2}{\kappa_{DE}}^{jk} + {\kappa_{HB}}^{jk};\ (\widetilde{\kappa}_{e-})^{jk} = \frac{1}{2}{\kappa_{DE}}^{jk} - {\kappa_{HB}}^{jk} - \frac{1}{3}\delta^{jk}(\kappa_{DE})^{ll} \\
(\widetilde{\kappa}_{o+})^{jk} &= \frac{1}{2}{\kappa_{DB}}^{jk} + {\kappa_{HE}}^{jk};\ \widetilde{\kappa}_{o-}^{jk} = \frac{1}{2}{\kappa_{DB}}^{jk} - {\kappa_{HE}}^{jk} \\
\widetilde{\kappa}_{tr} &= \frac{1}{3}\delta^{jk}(\kappa_{DE})^{ll}
\end{aligned}
\tag{14}
$$

In the above definitions, $\widetilde{\kappa}_{e+}$, $\widetilde{\kappa}_{e-}$ and $\widetilde{\kappa}_{tr}$ are parity-even matrices, while $\widetilde{\kappa}_{o+}$ and $\widetilde{\kappa}_{o-}$ are the parity-odd matrices. As mentioned in [7, 23], the $\widetilde{\kappa}_{e+}$ and $\widetilde{\kappa}_{o-}$ are constrained such that $\left|\widetilde{\kappa}^{JK}\right| \leq 2 \times 10^{-32}$. Thus, both $\widetilde{\kappa}_{e+}$ and $\widetilde{\kappa}_{o-}$ are set to zero each time they appear in our equations. Also, $\widetilde{\kappa}_{tr}$ is a scalar, which resonator experiments are usually insensitive to [24], and is not considered in this work.

The standard reference frame that we use is the Sun-Centered Celestial Equatorial Frame (SCCEF), which is shown in Fig. 2. This is the frame in which the sun is at the center, and is inertial with respect to the CMB to first order. The axes in this frame are labeled X, Y and Z. The Z axis is defined [7, p. 6] [25, p. 3] to be parallel to the Earth's north pole, or 90° declination. The X axis points from the sun toward the Earth at the moment of the autumn equinox, or 0° right ascension (RA) and 0° declination, while the Y axis is at 90° RA and also at 0° declination, usually taken in the J2000.0 frame.

The convention described in [7, p. 18], which has the raised capital indices (J, K) in the SCCEF, has been used. Local coordinates x, y and z are defined on the Earth's surface (at the point of the experiment). The z axis is defined as being locally normal to the ground, vertically upwards. The x axis points south and the y axis points east. These coordinates are denoted by the lowered capital indices (j, k) and they rotate with sidereal period $\Delta T_{\oplus} = \frac{1}{\omega_{\oplus}} \approx 23$ h 56 min. There is a relation between these two coordinates which is given by the following rotation matrix:

$$R^{jJ} = \begin{pmatrix} \cos\chi\cos\omega_\oplus T_\oplus & \cos\chi\sin\omega_\oplus T_\oplus & -\sin\chi \\ -\sin\omega_\oplus T_\oplus & \cos\omega_\oplus T_\oplus & 0 \\ \sin\chi\cos\omega_\oplus T_\oplus & \sin\chi\sin\omega_\oplus T_\oplus & \sin\chi \end{pmatrix} \tag{15}$$

Here χ is the co-latitude of the laboratory from the north pole, and the $T_\oplus$ is the time coordinate that is related to the sidereal frequency of the Earth. The time $T_\oplus$ is defined in [7] as any time when the y axis and the Y axis align. This has been taken to be the first time this occurs after the vernal equinox, which points along the negative x-axis.

When searching for leading order violations it is only necessary to consider the spinning of the Earth about itself. However, the orbit of the Earth about the sun may also be considered, since it induces Lorentz boosts and we may calculate proportional terms to these. Since the Earth moves relatively slowly around the Sun compared to the speed of light, the boost terms will be suppressed by the velocity with respect to the speed of light ($\beta_\oplus = \frac{v_\oplus}{c} \approx 10^{-4}$). The boost velocity of a point on the Earth's surface is given by the following relation:

$$\boldsymbol{\beta} = \beta_\oplus \begin{pmatrix} \sin\Omega_\oplus T \\ -\cos\eta\cos\Omega_\oplus T \\ -\sin\eta\cos\Omega_\oplus T \end{pmatrix} + \beta_L \begin{pmatrix} -\sin\omega_\oplus T \\ \cos\omega_\oplus T \\ 0 \end{pmatrix} \tag{16}$$

Here $\beta_\oplus$ is the value for the boost speed of the orbital motion of the Earth and β_L is the boost speed of the lab at the surface of the Earth due to its spin motion. The latter is location dependent, but is less than 1.5×10^{-6} and is zero at the poles (η is as defined in Fig. 2). The Lorentz matrix, Λ^μ_ν, that implements the transformation from the SCCEF to the laboratory frame with the sidereal rotation R^{jJ} and a boost $\boldsymbol{\beta}$ is given by,

$$\Lambda^\mu_\nu = \begin{pmatrix} 1 & -\beta^1 & -\beta^2 & -\beta^3 \\ -(R\cdot\boldsymbol{\beta})^1 & R^{11} & R^{12} & R^{13} \\ -(R\cdot\boldsymbol{\beta})^2 & R^{21} & R^{22} & R^{23} \\ -(R\cdot\boldsymbol{\beta})^3 & R^{31} & R^{32} & R^{33} \end{pmatrix} \tag{17}$$

After some calculation [7] the κ matrices from the SCCEF (indexed by J and K) can be express in terms of the values in the laboratory frame (indexed by j and k).

$$(\kappa_{DE})^{jk}_{lab} = T_0^{jkJK}(\kappa_{DE})^{JK} - T_1^{kjJK}(\kappa_{DB})^{JK} - T_1^{jkJK}(\kappa_{DB})^{JK} \tag{18}$$

$$(\kappa_{HB})^{jk}_{lab} = T_0^{jkJK}(\kappa_{DE})^{JK} - T_1^{(kjKJ}(\kappa_{DB})^{JK} - T_1^{jkKJ}(\kappa_{DB})^{JK} \tag{19}$$

$$(\kappa_{DB})^{jk}_{lab} = T_0^{jkJK}(\kappa_{DB})^{JK} + T_1^{kjJK}(\kappa_{DB})^{JK} + T_1^{jkJK}(\kappa_{HB})^{JK} \tag{20}$$

Where:

$$T_0^{jkJK} = R^{jJ}R^{kK} \tag{21}$$

$$T_1^{jkJK} = R^{jP}R^{kJ}\epsilon^{KPQ}\beta^Q \tag{22}$$

Here ϵ is the standard anti-symmetric tensor.

In the Sect. 3 we apply the above to calculate the sensitivity of typical resonator experiment. To do this the sensitivity to the components given in (12) are derived.

3 Applying the SME to Resonator Experiments

The modified Lagrangian of the SME introduces perturbations of the electric and magnetic fields in a vacuum. The unperturbed fields are denoted by a zero subscript to distinguish them from the Lorentz-violating fields. Putative Lorentz violations are produced by motion with respect to a preferred frame, which perturbs the fields generating an observable signal. The general framework [7] for denoting the sensitivity of this observable signal in the laboratory frame is a linear expression as follows:

$$\delta\mathcal{O} = (\mathcal{M}_{DE})^{jk}_{lab}(\kappa_{DE})^{jk}_{lab} + (\mathcal{M}_{HB})^{jk}_{lab}(\kappa_{HB})^{jk}_{lab} + (\mathcal{M}_{DB})^{jk}_{lab}(\kappa_{DB})^{jk}_{lab} \tag{23}$$

The summation over the indices is implied, and the components of the $\mathcal{M}^{jk}_{lab}$ matrices are in general a function of time. The observable is dependent on the type of experiment, and in the case of a resonant cavity experiments it is the resonance frequency deviation, $\frac{\delta\nu}{\nu_0}$. Since the laboratory frame and the resonator frame do not necessarily coincide, we first consider $\mathcal{M}^{jk}_{res}$ coefficients in the resonator frame and later relate it to the laboratory and sun-centered frame.

In general, resonators may be constructed from dielectric and magnetic materials. To calculate the $\mathcal{M}^{jk}_{res}$ matrices for such structures a more general form of (13) must be considered, which includes the properties of the medium, μ_r (permeability) and ϵ_r (permittivity), which are in general second order tensors.

$$\begin{pmatrix} \mathbf{D} \\ \mathbf{H} \end{pmatrix} = \begin{pmatrix} \epsilon_0(\epsilon_r + \kappa_{DE}) & \sqrt{\frac{\epsilon_0}{\mu_0}}\kappa_{DB} \\ \sqrt{\frac{\epsilon_0}{\mu_0}}\kappa_{HE} & \mu_0^{-1}(\mu_r^{-1} + \kappa_{HB}) \end{pmatrix} \begin{pmatrix} \mathbf{E} \\ \mathbf{B} \end{pmatrix} \tag{24}$$

Here as was derived in [7], we assume that the fractional frequency shift due to Lorentz violations is given by:

$$\frac{\Delta\nu}{\nu_0} = -\frac{1}{4\langle U\rangle} \times \int_V d^3x \left(\epsilon_0 \mathbf{E}_0^* \cdot \kappa_{DE} \cdot \mathbf{E}_0 - \mu_0^{-1}\mathbf{B}_0^* \cdot \kappa_{HB} \cdot \mathbf{B}_0 + 2Re\left(\sqrt{\frac{\epsilon_0}{\mu_0}}\mathbf{E}_0^* \cdot \kappa_{DB} \cdot \mathbf{B}_0\right)\right) \tag{25}$$

Here $\langle U\rangle$ is the energy stored in the field and is given by the standard electrodynamic integral.

$$\langle U\rangle = \frac{1}{4}\int_V d^3x(\mathbf{E}_0 \cdot \mathbf{D}_0^* + \mathbf{B}_0 \cdot \mathbf{H}_0^*) \tag{26}$$

In Maxwellian electrodynamics the balance of magnetic and electrical energy in a resonator is equal, so the following is true:

$$\langle U\rangle = \frac{1}{2}\int_V d^3x \mathbf{E}_0 \cdot \mathbf{D}_0^* = \frac{1}{2}\int_V d^3x \mathbf{B}_0 \cdot \mathbf{H}_0^* \tag{27}$$

This reduces $\langle U\rangle$ to an effective normalisation factor for either an electric or magnetic filling factor. Also, since the κ terms from the integral of (25) are only

time dependent rather than spatially dependent, the κ terms can be removed from the integral. Thus, the final term in (25) will be zero since the electric and magnetic fields are orthogonal in a resonant structure. By equating (23) and (25) in the resonator frame the $(\mathcal{M}_{DB})^{jk}_{res}$ coefficients are calculated to be zero, eliminating the possibility of making a measurement of κ_{tr} [24, 26]. Assuming the resonator permeability and permittivity have no off-diagonal coefficients (i.e. non-gyrotropic) such that;

$$\epsilon_r = \epsilon_0 \begin{pmatrix} \epsilon_x & 0 & 0 \\ 0 & \epsilon_y & 0 \\ 0 & 0 & \epsilon_z \end{pmatrix} \quad \mu_r = \mu_0 \begin{pmatrix} \mu_x & 0 & 0 \\ 0 & \mu_y & 0 \\ 0 & 0 & \mu_z \end{pmatrix} \tag{28}$$

the only non-zero coefficients may then be calculated to be

$$(\mathcal{M}_{DE})^{jj}_{res} = -\frac{1}{\epsilon_j} \frac{\int_V d^3x \left| E_0^j \right|^2}{2\int_V d^3x \mathbf{E_0}^* \cdot \mathbf{E_0}} = -\frac{Pe_j}{2\epsilon_j} \tag{29}$$

$$(\mathcal{M}_{HB})^{jj}_{res} = \mu_j \frac{\int_V d^3x \left| H_0^j \right|^2}{2\int_V d^3x \mathbf{H_0}^* \cdot \mathbf{H_0}} = \mu_j \frac{Pm_j}{2} \tag{30}$$

Thus the $\mathcal{M}_{DE}$ and $\mathcal{M}_{HB}$ matrices are diagonal and simply related to the electric and magnetic energy filling factors, Pe_j and Pm_j respectively [11]. In general a resonator may consist of more than one material, and may include vacuum. In this case (29) and (30) may be written more generally (s is the number of different materials including vacuum).

$$(\mathcal{M}_{DE})^{jj}_{res} = -\sum_{i=1}^{s} \frac{Pe_j^i}{2\epsilon_j^i} \tag{31}$$

$$(\mathcal{M}_{HB})^{jj}_{res} = \sum_{i=1}^{s} \frac{\mu_j^i Pm_j^i}{2} \tag{32}$$

To measure the resonant frequency it is necessary excite electromagnetic fields inside the resonator and then compare it against a similar frequency. To be sensitive to violations of LLI, the comparison frequency must be generated by a source which exhibits a different dependence on Lorentz violations in the photon sector. For example, an atomic standard (such as a hydrogen maser) may operate in a mode which is not sensitive to Lorentz violations [11, 13]. Alternatively, the resonant frequency may be compared against another resonator designed to have a different dependence. The latter can be achieved by orientating two identical resonators orthogonally [10], or by exciting two modes in a matter filled resonator with orthogonal polarizations. In both cases the field components must be considered with respect to the laboratory frame and not the resonator. For such an experiment the observable becomes the frequency difference (between a resonator labeled a and b) such that;

$$\delta\mathcal{O} = \frac{\delta\nu_a}{\nu_a} - \frac{\delta\nu_b}{\nu_b} \tag{33}$$

Thus, with respect to the laboratory frame, the effective $(\mathcal{M}_{DE})_{lab}$ and $(\mathcal{M}_{HB})_{lab}$ matricies consistent with (23) become:

$$(\mathcal{M}_{DE})_{a-b} = \tag{34}$$
$$\begin{pmatrix} (\mathcal{M}_{DE})_a^{xx} - (\mathcal{M}_{DE})_b^{xx} & 0 & 0 \\ 0 & (\mathcal{M}_{DE})_a^{yy} - (\mathcal{M}_{DE})_b^{yy} & 0 \\ 0 & 0 & (\mathcal{M}_{DE})_a^{zz} - (\mathcal{M}_{DE})_b^{zz} \end{pmatrix}$$

$$(\mathcal{M}_{HB})_{a-b} = \tag{35}$$
$$\begin{pmatrix} (\mathcal{M}_{HB})_a^{xx} - (\mathcal{M}_{HB})_b^{xx} & 0 & 0 \\ 0 & (\mathcal{M}_{HB})_a^{yy} - (\mathcal{M}_{HB})_b^{yy} & 0 \\ 0 & 0 & (\mathcal{M}_{HB})_a^{zz} - (\mathcal{M}_{HB})_b^{zz} \end{pmatrix}$$

These equations are general for any resonator experiments, including Fabry-Perot and microwave cavity experiments, and simplify the analysis for complex resonator configurations, such as whispering gallery mode resonators. Only the electric and magnetic filling factors need to be calculated to determine the sensitivity coefficients to the observable, which is possible using standard numerical techniques [27].

To determine the sensitivity of stationary laboratory experiments one calculates the time dependence of (33) due to the sidereal and orbital motion of the Earth around the Sun in terms of the Sun-centered coefficients given in (14) and (18). This calculation has already been done in [7, 11, 28] and will not be repeated here. In the following subsection we generalize this analysis to rotating experiments.

3.1 Rotation in the Laboratory Frame

Non-rotating experiments [10, 12, 13] that rely on Earth rotation to modulate a Lorentz violating effect are not optimal for two reasons. Firstly, the sensitivity is proportional to the noise in the system at the modulation frequency, typically best for microwave resonator-oscillators and Fabry-Perot stabilized lasers for periods between 10 to 100 seconds. Secondly, the sensitivity is proportional to the square root of the number of periods of the modulation signal, therefore taking a relatively long time to acquire sufficient data. Thus, by rotating the experiment the data integration rate is increased and the relevant signals are translated to the optimal operating regime [14]. For rotation in the laboratory frame the $(\mathcal{M})_{lab}^{jk}$ coefficients become a function of time and depend on the axis of rotation. In the laboratory it is most practical to rotate around the axis of the gravitational field to reduce gravity induced perturbation of the experiment. Thus, our analysis includes rotation about the laboratory z-axis. If we set the

time, $t = 0$, to be defined when the experiment and laboratory axes are aligned, and we only consider the time varying components (i.e. the most sensitive ones induced by rotation), then for clock-wise rotation of ω_R rad/s, (34) and (35) become:

$$(\mathcal{M}_{DE})_{lab} = \begin{pmatrix} \mathcal{S}_{DE}\cos(2\omega_R t) & -\mathcal{S}_{DE}\sin(2\omega_R t) & 0 \\ -\mathcal{S}_{DE}\sin(2\omega_R t) & -\mathcal{S}_{DE}\cos(2\omega_R t) & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{36}$$

$$\mathcal{S}_{DE} = \frac{1}{2}\left((\mathcal{M}_{DE})_a^{xx} - (\mathcal{M}_{DE})_a^{yy} - (\mathcal{M}_{DE})_b^{xx} + (\mathcal{M}_{DE})_b^{yy}\right) \tag{37}$$

$$(\mathcal{M}_{HB})_{lab} = \begin{pmatrix} \mathcal{S}_{HB}\cos(2\omega_R t) & -\mathcal{S}_{HB}\sin(2\omega_R t) & 0 \\ -\mathcal{S}_{HB}\sin(2\omega_R t) & -\mathcal{S}_{HB}\cos(2\omega_R t) & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{38}$$

$$\mathcal{S}_{HB} = \frac{1}{2}\left((\mathcal{M}_{HB})_a^{xx} - (\mathcal{M}_{HB})_a^{yy} - (\mathcal{M}_{HB})_b^{xx} + (\mathcal{M}_{HB})_b^{yy}\right) \tag{39}$$

Note that if one resonator is tested with respect to a stationary generated frequency, then the $(\mathcal{M})_i^{jj}$ coefficients in the definition of $\mathcal{S}_{HB}$ and $\mathcal{S}_{HB}$ pertaining to that frequency must be set to zero.

To determine the time dependence of the observable (33) we follow the same procedure as presented in Subsect. 2.2 to transform the $\tilde{\kappa}$ matricies given in (14) to the κ matricies in the laboratory given by (18), (19) and (20). We then substitute (37) to (39) into (23) to calculate the time dependence of the observable of (33). This is a tedious process and the details are omitted. Essentially, because the $(\mathcal{M}_{DE})_{lab}$ and $(\mathcal{M}_{HB})_{lab}$ matrices are time dependent at $2\omega_R$, the observable signals are at frequencies close to this value and are summarized in Table 1. Here the frequency of Earth rotation is defined as $\omega_\oplus$, and orbit around the sun as $\Omega_\oplus$. The $\omega_\oplus$ is commonly referred to as the sidereal frequency, while the $\Omega_\oplus$ is referred to as the annual frequency. We also define the sensitivity factor, $\mathcal{S}$ of the experiment as:

$$\mathcal{S} = \mathcal{S}_{HB} - \mathcal{S}_{DE} \tag{40}$$

To decorrelate all side bands, more than one year of data is necessary. In this case we have eight unknown $\tilde{\kappa}$ coefficients and thirty possible individual measurements listed in Table 1, which is an over parameterization. For short data sets (less than a year) we do not have enough information to satisfy the Nyquist condition to distinguish between frequencies that differ by the annual offset (collected in the same blocks). Thus, to make a short data set approximation, we collect the sidebands together (see Fig. 3). The short data set approximation is achieved by knowing the angle of the orbit, $\Phi = \Omega_\oplus t$, in the sun-centered frame with respect to the negative X-axis (which occurs at the vernal equinox as shown in Fig. 2), and then taking a Taylor series expansion around that angle. Here we define the phase of the combined rotational and sidereal term as θ and Φ_0 as the value of Φ when a short data set is taken. Since $\delta\Phi \equiv \Phi - \Phi_0$ is small with respect to 2π, via the double angle rule we can derive the following relationships:

Table 1. Normalized sensitivities with respect to the experiment sensitivity factor $\mathcal{S}$ for all predicted frequency modulated components

ω_i	Cosine Coefficient $C_{\omega_i}/\mathcal{S}$	Sine Coefficient $S_{\omega_i}/\mathcal{S}$
$2\omega_R$	$\frac{3}{2}\sin^2(\chi)\widetilde{\kappa}_{e-}^{ZZ}$	$2\beta_L\sin(\chi)\widetilde{\kappa}_{o+}^{XY}$
$2\omega_R+\Omega_\oplus$	$-\frac{1}{2}\beta_\oplus\sin^2(\chi)\times$ $(2\sin(\eta)\widetilde{\kappa}_{o+}^{XY}+\cos(\eta)\widetilde{\kappa}_{o+}^{XZ})$	$-\frac{1}{2}\beta_\oplus\sin^2(\chi)\widetilde{\kappa}_{o+}^{YZ}$
$2\omega_R-\Omega_\oplus$	$-\frac{1}{2}\beta_\oplus\sin^2(\chi)\times$ $(2\sin(\eta)\widetilde{\kappa}_{o+}^{XY}+\cos(\eta)\widetilde{\kappa}_{o+}^{XZ})$	$\frac{1}{2}\beta_\oplus\sin^2(\chi)\widetilde{\kappa}_{o+}^{YZ}$
$2\omega_R+\omega_\oplus$	$-2\sin^2(\frac{\chi}{2})\times$ $(-\beta_L\widetilde{\kappa}_{o+}^{XZ}+\sin(\chi)\widetilde{\kappa}_{e-}^{XZ})$	$-2\sin^2(\frac{\chi}{2})(\sin(\chi)\widetilde{\kappa}_{e-}^{YZ}-\beta_L\widetilde{\kappa}_{o+}^{YZ})$
$2\omega_R+\omega_\oplus+\Omega_\oplus$	$-2\beta_\oplus\cos(\frac{\chi}{2})\sin(\eta)\sin^3(\frac{\chi}{2})\widetilde{\kappa}_{o+}^{YZ}$	$4\beta_\oplus\cos(\frac{\chi}{2})\sin(\frac{\eta}{2})\sin^3(\frac{\chi}{2})\times$ $(\sin(\frac{\eta}{2})\widetilde{\kappa}_{o+}^{XY}+\cos(\frac{\eta}{2})\widetilde{\kappa}_{o+}^{XZ})$
$2\omega_R+\omega_\oplus-\Omega_\oplus$	$-2\beta_\oplus\cos(\frac{\chi}{2})\sin(\eta)\sin^3(\frac{\chi}{2})\widetilde{\kappa}_{o+}^{YZ}$	$-4\beta_\oplus\cos(\frac{\chi}{2})\cos(\frac{\eta}{2})\sin^3(\frac{\chi}{2})\times$ $(\cos(\frac{\eta}{2})\widetilde{\kappa}_{o+}^{XY}-\sin(\frac{\eta}{2})\widetilde{\kappa}_{o+}^{XZ})$
$2\omega_R+2\omega_\oplus$	$-\sin^4(\frac{\chi}{2})(\widetilde{\kappa}_{e-}^{XX}-\widetilde{\kappa}_{e-}^{YY})$	$-2\sin^4(\frac{\chi}{2})\widetilde{\kappa}_{e-}^{XY}$
$2\omega_R+2\omega_\oplus+\Omega_\oplus$	$2\beta_\oplus\sin^2(\frac{\eta}{2})\sin^4(\frac{\chi}{2})\widetilde{\kappa}_{o+}^{XZ}$	$2\beta_\oplus\sin^2(\frac{\eta}{2})\sin^4(\frac{\chi}{2})\widetilde{\kappa}_{o+}^{YZ}$
$2\omega_R+2\omega_\oplus-\Omega_\oplus$	$-2\beta_\oplus\cos^2(\frac{\eta}{2})\sin^4(\frac{\chi}{2})\widetilde{\kappa}_{o+}^{XZ}$	$-2\beta_\oplus\cos^2(\frac{\eta}{2})\sin^4(\frac{\chi}{2})\widetilde{\kappa}_{o+}^{YZ}$
$2\omega_R-\omega_\oplus$	$2\cos^2(\frac{\chi}{2})(\beta_L\widetilde{\kappa}_{o+}^{XZ}+\sin(\chi)\widetilde{\kappa}_{e-}^{XZ})$	$-2\cos^2(\frac{\chi}{2})(\beta_L\widetilde{\kappa}_{o+}^{YZ}+\sin(\chi)\widetilde{\kappa}_{e-}^{YZ})$
$2\omega_R-\omega_\oplus+\Omega_\oplus$	$2\beta_\oplus\cos^3(\frac{\chi}{2})\sin(\eta)\sin(\frac{\chi}{2})\widetilde{\kappa}_{o+}^{YZ}$	$-4\beta_\oplus\cos(\frac{\eta}{2})\cos^3(\frac{\chi}{2})\sin(\frac{\chi}{2})\times$ $(\cos(\frac{\eta}{2})\widetilde{\kappa}_{o+}^{XY}-\sin(\frac{\eta}{2})\widetilde{\kappa}_{o+}^{XZ})$
$2\omega_R-\omega_\oplus-\Omega_\oplus$	$2\beta_\oplus\cos^3(\frac{\chi}{2})\sin(\eta)\sin(\frac{\chi}{2})\widetilde{\kappa}_{o+}^{YZ}$	$4\beta_\oplus\sin(\frac{\eta}{2})\cos^3(\frac{\chi}{2})\sin(\frac{\chi}{2})\times$ $(\sin(\frac{\eta}{2})\widetilde{\kappa}_{o+}^{XY}+\cos(\frac{\eta}{2})\widetilde{\kappa}_{o+}^{XZ})$
$2\omega_R-2\omega_\oplus$	$-\cos^4(\frac{\chi}{2})(\widetilde{\kappa}_{e-}^{XX}-\widetilde{\kappa}_{e-}^{22})$	$2cos^4(\frac{\chi}{2})\widetilde{\kappa}_{e-}^{XY}$
$2\omega_R-2\omega_\oplus+\Omega_\oplus$	$-2\beta_\oplus\cos^2(\frac{\eta}{2})\cos^4(\frac{\chi}{2})\widetilde{\kappa}_{o+}^{XZ}$	$2\beta_\oplus\cos^2(\frac{\eta}{2})\cos^4(\frac{\chi}{2})\widetilde{\kappa}_{o+}^{YZ}$
$2\omega_R-2\omega_\oplus-\Omega_\oplus$	$2\beta_\oplus\sin^2(\frac{\eta}{2})\cos^4(\frac{\chi}{2})\widetilde{\kappa}_{o+}^{XZ}$	$-2\beta_\oplus\sin^2(\frac{\eta}{2})\cos^4(\frac{\chi}{2})\widetilde{\kappa}_{o+}^{YZ}$

$$\begin{aligned}\sin(\theta\pm(\delta\Phi+\Phi_0)) &= \sin(\theta\pm\delta\Phi)\cos(\Phi_0)\pm\cos(\theta\pm\delta\Phi)\sin(\Phi_0)\\ &\approx \sin(\theta)\cos(\Phi_0)\pm\cos(\theta)\sin(\Phi_0)\\ \cos(\theta\pm(\delta\Phi+\Phi_0)) &= \cos(\theta\pm\delta\Phi)\cos(\Phi_0)\mp\sin(\theta\pm\delta\Phi)\sin(\Phi_0)\\ &\approx \cos(\theta)\cos(\Phi_0)\mp\sin(\theta)\sin(\Phi_0)\end{aligned}$$

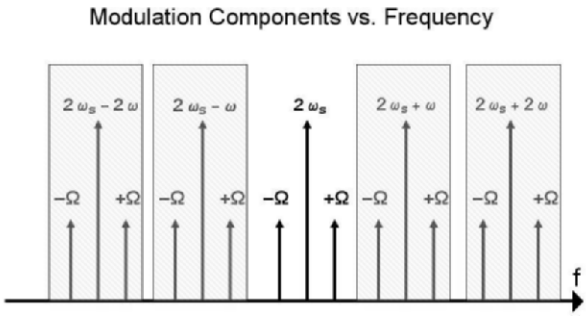


Fig. 3. This "frequency stick diagram" shows a schematic of the frequency modulation components of the beat frequency in a convenient form. The sidebands offset by $\pm\Omega_\oplus$ have been trimmed for brevity. Each frequency has two degrees of freedom, i.e. a sine and a cosine term, with the phase set by a time $t = 0$ set with respect to the SCCEF

Now we can combine the sidebands as shown in Fig. 3 by applying the above relationships to eliminate the dependence on $\Omega_\oplus$. In this case the components from Table 1 decompose to those listed in Table 2.

The first feature to notice is that the $2\omega_R \pm 2\omega_\oplus$ sidebands are redundant. One might also expect the $2\omega_R \pm \omega_\oplus$ sidebands to be redundant as well. The only reason they are not is because we have taken into account the velocity of the laboratory due to the Earth spinning on its axis, β_L. In fact it turns out that it is not useful to keep this term because $\beta_\oplus$ is two orders of magnitude larger and when one applies the data analysis procedures the sensitivities will be degraded if the analysis depends on the β_L terms for the uniqueness of the solution. Thus, since it makes no practical sense to keep these terms, we set them to zero. For this case the coefficients are listed in Table 3.

For data sets of less than one year, the components in Table 3 may be used to set upper limits on the $\tilde{\kappa}$ coefficients in the SME. Since there are only five possible independent components, to set limits on eight coefficients we use the same technique as adopted by Lipa et al. [12]. The $\tilde{\kappa}_{o+}$ boost coefficients are set to zero to calculate limits on the $\tilde{\kappa}_{e-}$ isotropy coefficients and vice versa. This technique assumes no correlation between the isotropy and boost coefficients. It would be unlikely that a cancellation of Lorentz violating effects would occur, as this would necessitate a fortuitous relationship between the coefficients of the same order of value as the boost suppression coefficient (i.e. orbit velocity, $\beta_\oplus$), and consistent with the correct linear combinations as presented in Table 3.

Another practical point is that the largest systematic effect occurs at $2\omega_R$. Thus, when setting the limits on the three $\widetilde{\kappa}_{o+}^{XY}$, $\widetilde{\kappa}_{o+}^{XZ}$ and $\widetilde{\kappa}_{o+}^{YZ}$ coefficients we only use the data collected at $2\omega_R \pm \omega_\oplus$ and $2\omega_R \pm 2\omega_\oplus$ frequencies. Likewise for the $\widetilde{\kappa}_{e-}^{XY}$, $\widetilde{\kappa}_{e-}^{XZ}$, $\widetilde{\kappa}_{e-}^{YZ}$ and $(\widetilde{\kappa}_{e-}^{XX} - \widetilde{\kappa}_{e-}^{YY})$ coefficients. These are the same coefficients that have had limits set by the non-rotating experiments [10,12,13]. The remaining coefficient $\widetilde{\kappa}_{e-}^{ZZ} (\equiv \widetilde{\kappa}_{e-}^{XX} + \widetilde{\kappa}_{e-}^{YY})$ can only be set amongst a systematic signal at $2\omega_R$, which is in general much greater than the statistical uncertain-

Table 2. Normalized sensitivities with respect to the experiment sensitivity factor $\mathcal{S}$ for all predicted frequency modulated components using the short data set approximation

	Coefficient (normalized with respect to $\mathcal{S}$)
$S_{2\omega_R-2\omega_\oplus}$	$-\cot^4(\frac{\chi}{2}) \times S_{2\omega_R+2\omega_\oplus}$
$C_{2\omega_R-2\omega_\oplus}$	$\cot^4(\frac{\chi}{2}) \times C_{2\omega_R+2\omega_\oplus}$
$S_{2\omega_R-\omega_\oplus}$	$-2\cos^2(\frac{\chi}{2})\times$ $(\beta_L\widetilde{\kappa}_{o+}^{YZ} + \sin(\chi)(\beta_\oplus\cos(\Phi_0)(\cos(\eta)\widetilde{\kappa}_{o+}^{XY} - \sin(\eta)\widetilde{\kappa}_{o+}^{XZ}) + \widetilde{\kappa}_{e-}^{YZ}))$
$C_{2\omega_R-\omega_\oplus}$	$2\cos^2(\frac{\chi}{2})\times$ $(\beta_L\widetilde{\kappa}_{o+}^{XZ} + \sin(\chi)(-\beta_\oplus\sin(\Phi_0)\widetilde{\kappa}_{o+}^{XY} + \beta_\oplus\cos(\Phi_0)\sin(\eta)\widetilde{\kappa}_{o+}^{YZ} + \widetilde{\kappa}_{e-}^{XZ}))$
$S_{2\omega_R}$	$2\beta_L\sin(\chi)\widetilde{\kappa}_{o+}^{XY}$
$C_{2\omega_R}$	$-\frac{1}{2}\sin^2(\chi)(2\beta_\oplus\cos(\Phi_0)(2\sin(\eta)\widetilde{\kappa}_{o+}^{XY} + \cos(\eta)\widetilde{\kappa}_{o+}^{XZ}) + 2\beta_\oplus\sin(\Phi_0)\widetilde{\kappa}_{o+}^{YZ})$ $+\frac{3}{2}\sin^2(\chi)\widetilde{\kappa}_{e-}^{ZZ})$
$S_{2\omega_R+\omega_\oplus}$	$-2\sin^2(\frac{\chi}{2})\times$ $(-\beta_L\widetilde{\kappa}_{o+}^{YZ} + \sin(\chi)(\beta_\oplus\cos(\Phi_0)(\cos(\eta)\widetilde{\kappa}_{o+}^{XY} - \sin(\eta)\widetilde{\kappa}_{o+}^{XZ}) + \widetilde{\kappa}_{e-}^{YZ}))$
$C_{2\omega_R+\omega_\oplus}$	$-2\sin^2(\frac{\chi}{2})\times$ $(-\beta_L\widetilde{\kappa}_{o+}^{XZ} + \sin(\chi)(-\beta_\oplus\sin(\Phi_0)\widetilde{\kappa}_{o+}^{XY} + \beta_\oplus\cos(\Phi_0)\sin(\eta)\widetilde{\kappa}_{o+}^{YZ} + \widetilde{\kappa}_{e-}^{XZ}))$
$S_{2\omega_R+2\omega_\oplus}$	$-2\sin^4(\frac{\chi}{2})(\beta_\oplus\sin(\Phi_0)\widetilde{\kappa}_{o+}^{XZ} + \beta_\oplus\cos(\eta)\cos(\Phi_0)\widetilde{\kappa}_{o+}^{YZ} + \widetilde{\kappa}_{e-}^{XY})$
$C_{2\omega_R+2\omega_\oplus}$	$-\sin^4(\frac{\chi}{2})(2\beta_\oplus\cos(\eta)\cos(\Phi_0)\widetilde{\kappa}_{o+}^{XZ} - 2\beta_\oplus\sin(\Phi_0)\widetilde{\kappa}_{o+}^{YZ} + (\widetilde{\kappa}_{e-}^{XX} - \widetilde{\kappa}_{e-}^{YY}))$

ties at the other frequencies. In this case we can assume all coefficients are zero except for the $\widetilde{\kappa}_{e-}^{ZZ}$ coefficient. However, it is not straight forward to set a limit on any putative Lorentz violation amongst a large systematic as one can not be sure if the systematic signal actually cancels an effect. Since the signal at $2\omega_R$ is dominated by systematic effects, it is likely that its phase and amplitude will vary across different data sets. In this case the systematic signal from multiple data sets can be treated statistically to place an upper limit on $\widetilde{\kappa}_{e-}^{ZZ}$. In our experiment we use this technique to set an upper of $2.1(5.7) \times 10^{-14}$ [15] (see Sect. 6).

3.2 Phase with Respect to the SCCEF

To extract the κ components of the SME out of our observed signal we first need to determine the relevant $C_{2\omega_i}$ and $S_{2\omega_i}$ coefficients listed in Table 1. This in turn requires us to know the phase of the experiment's orientation with respect to the SCCEF. In this section we will derive an expression for this phase in terms of the time origins of the experiment's rotation, the Earth's sidereal rotation, and the orbit of the Earth around the Sun.

In general, we are interested in the frequency components

Table 3. Normalized sensitivities with respect to the experiment sensitivity factor $\mathcal{S}$ for all predicted frequency modulated components, using the short data set approximation and neglecting components of order β_L

	Coefficient (normalized with respect to $\mathcal{S}$)
$S_{2\omega_R-2\omega_\oplus}$	$-\cot^4(\frac{\chi}{2}) \times S_{2\omega_R+2\omega_\oplus}$
$C_{2\omega_R-2\omega_\oplus}$	$\cot^4(\frac{\chi}{2}) \times C_{2\omega_R+2\omega_\oplus}$
$S_{2\omega_R-\omega_\oplus}$	$\cot^2(\frac{\chi}{2}) \times S_{2\omega_R+\omega_\oplus}$
$C_{2\omega_R-\omega_\oplus}$	$-\cot^2(\frac{\chi}{2}) \times C_{2\omega_R+\omega_\oplus}$
$S_{2\omega_R}$	—-
$C_{2\omega_R}$	$-\frac{1}{2}\sin^2(\chi)(2\beta_\oplus\cos(\Phi_0)(2\sin(\eta)\widetilde{\kappa}_{o+}^{XY}+\cos(\eta)\widetilde{\kappa}_{o+}^{XZ})+2\beta_\oplus\sin(\Phi_0)\widetilde{\kappa}_{o+}^{YZ})$ $+\frac{3}{2}\sin^2(\chi)\widetilde{\kappa}_{e-}^{ZZ})$
$S_{2\omega_R+\omega_\oplus}$	$-2\sin^2(\frac{\chi}{2})(\sin(\chi)(\beta_\oplus\cos(\Phi_0)(\cos(\eta)\widetilde{\kappa}_{o+}^{XY}-\sin(\eta)\widetilde{\kappa}_{o+}^{XZ})+\widetilde{\kappa}_{e-}^{YZ}))$
$C_{2\omega_R+\omega_\oplus}$	$-2\sin^2(\frac{\chi}{2})(\sin(\chi)(-\beta_\oplus\sin(\Phi_0)\widetilde{\kappa}_{o+}^{XY}+\beta_\oplus\cos(\Phi_0)\sin(\eta)\widetilde{\kappa}_{o+}^{YZ}+\widetilde{\kappa}_{e-}^{XZ}))$
$S_{2\omega_R+2\omega_\oplus}$	$-2\sin^4(\frac{\chi}{2})(\beta_\oplus\sin(\Phi_0)\widetilde{\kappa}_{o+}^{XZ}+\beta_\oplus\cos(\eta)\cos(\Phi_0)\widetilde{\kappa}_{o+}^{YZ}+\widetilde{\kappa}_{e-}^{XY})$
$C_{2\omega_R+2\omega_\oplus}$	$-\sin^4(\frac{\chi}{2})(2\beta_\oplus\cos(\eta)\cos(\Phi_0)\widetilde{\kappa}_{o+}^{XZ}-2\beta_\oplus\sin(\Phi_0)\widetilde{\kappa}_{o+}^{YZ}+(\widetilde{\kappa}_{e-}^{XX}-\widetilde{\kappa}_{e-}^{YY}))$

$$2\omega_{[a,b]} = 2\omega_R + a\omega_\oplus + b\Omega_\oplus \tag{41}$$

where a and b take on values in the domains

$$a \in [-2, 2], b \in [-1, 1] \tag{42}$$

Thus to determine the $C_{2\omega_{[a,b]}}$ coefficient we fit the data with a model of the form

$$\cos(2\omega_R T_R + a\omega_\oplus T_\oplus + b\Omega_\oplus T) \tag{43}$$

where T_R is the experiment's rotation time, $T_\oplus$ is the sidereal time, and T is the time since the vernal equinox.

To simplify our analysis we aim to transform this expression to the form $\cos(\alpha t + \phi)$. To achieve this we note that the difference δ_R between the experiment's rotation time T_R and the time since the vernal equinox T is constant over the course of the measurement, as determined by the initial configuration of the experiment, and we may write,

$$\delta_R = T_R - T. \tag{44}$$

Similarly the sidereal time and the time since the vernal equinox are related by $\delta_\oplus$,

$$\delta_\oplus = T_\oplus - T \tag{45}$$

By combining (43), (44) and (45) we arrive at an expression of the desired form.

$$\begin{aligned}\cos(2\omega_R T_R + a\omega_\oplus T_\oplus + b\Omega_\oplus T) &= \cos(2\omega_R(\delta_R + T) + a\omega_\oplus(\delta_\oplus + T) + b\Omega_\oplus T) \\ &= \cos((2\omega_R + a\omega_\oplus + b\Omega_\oplus)T \\ &\quad +2\omega_R\delta_R + a\omega_\oplus\delta_\oplus)\end{aligned}$$

Thus we can account for the phase of the experiment relative to the SCCEF by determining δ_R and $\delta_\oplus$. The origin of the experiment's rotation time T_R is defined to be the instant at which the axis of symmetry of the first resonator (resonator a) is aligned with the local y axis. Our experiment has been designed such that the time origin of the data acquisition coincides with the same event, rendering $\delta_R = 0$ in our case.

We also need to obtain $\delta_\oplus$ for the sidereal rotation. We define $T_\oplus = 0$ as in [7] to be the instant the local y axis and the SCCEF Y axis are aligned (noon) in the laboratory (see Fig. 2). Let us define T_v to be the time in seconds after midnight UTC+0, at which the vernal equinox has occurred in the J2000.0 frame [7]. For convenience we also define our longitude T_L in terms of sidereal seconds from midnight (in the case of our laboratory $T_l = 115.826^\circ \times \frac{23\text{ h }56\text{ min}}{360^\circ} = 27721$ s). There exists a special location whose meridian is at noon at the vernal equinox. For this special location (during the vernal equinox), $\delta_\oplus = 0$ since the time when the y and Y axes align and the vernal equinox are the same. We see geometrically that any longitude greater than this meridian will have positive $\delta_\oplus$, otherwise if the longitude is less than this meridian it would have negative $\delta_\oplus$. As shown in Fig. 4, we can now derive an expression for $\delta_\oplus$.

$$\delta_\oplus = T_L + T_v - \frac{23\text{ h }56\text{ min}}{2} \tag{46}$$

Hence we are able to determine the phase of the experiment's orientation relative to the SCCEF.

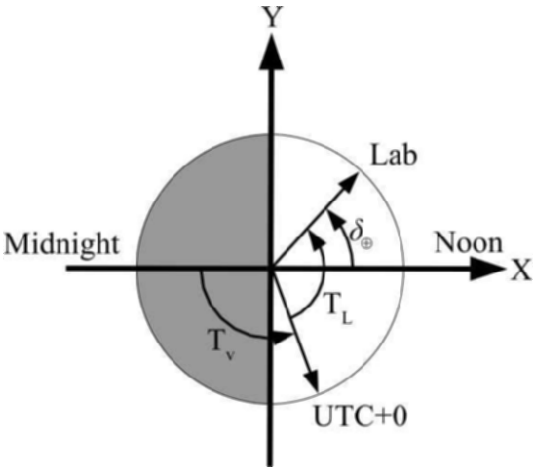


Fig. 4. Diagram showing meridians and angles

4 Comparison of Sensitivity of Various Resonator Experiments in the SME

In this section we show how the general analysis may be applied to some common resonator configurations for testing LLI. Also, we propose some new configurations based on exciting two modes in matter filled resonators. The comparison is made by calculating the sensitivity parameter $\mathcal{S}$ of the resonator using (31) to (40). Note that the sign of the $\mathcal{S}$ factor depends on the definition of the first resonator. Practically this will need to be the resonator that exhibits the largest value of frequency. In this work, where appropriate, we assume the first resonator is aligned along the y-axis.

4.1 Fabry-Perot Resonators

Experiments based on laser stabilized Fabry-Perot resonators typically use either one [8] or two [10] cavities placed with the lengths orthogonal to the laboratory z-axis. In a vacuum filled cavity it is easy to show that $|\mathcal{S}| = \frac{1}{2}$ for the configuration in Fig. 5. In contrast, when one rotating cavity is compared to a stationary one the value is reduced by a factor of 2, to $|\mathcal{S}| = \frac{1}{4}$. It is also interesting to consider the sensitivity of matter filled cavities in the photon sector. Here, for simplicity we assume the relative permeability and permittivity are scalars of μ_r and ϵ_r respectively. It is straight forward to add anisotropy and only modifies the sensitivity slightly, so for brevity is not considered here. If similar configurations to Fig. 5 are constructed from solid material the sensitivity factor, $\mathcal{S}$, becomes dependent on polarization. This effect also allows for a sensitive experiment by exciting two modes of different polarization inside one cavity (Dual-Mode), of which some examples are shown in Fig . 6. Such cavities have been built previously to measure birefringence [29]. High finesse matter cavities can be made by using low-loss crystalline dielectric materials at optical frequencies [30, 31]. The sensitivity for different configurations are compared in Table 4. For a low-loss dielectric material with E_z polarization in the two orthogonal cavities (Fig. 5) the

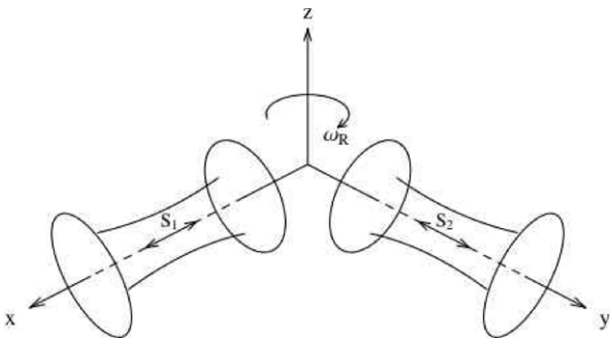


Fig. 5. Typical configuration of a rotating Lorentz invariance test using Fabry-Perot cavities

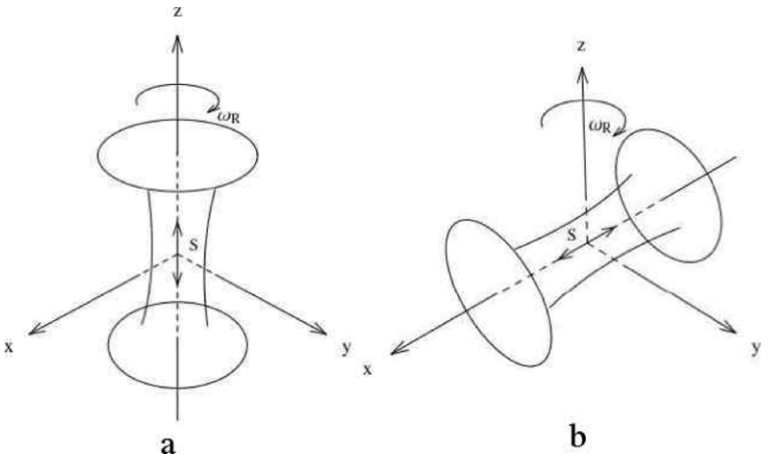


Fig. 6. New proposed matter filled Fabry-Perot cavity configurations in which two modes of orthogonal polarization are excited and compared, and are sensitive to violations in Lorentz invariance in the photon sector of the SME

Table 4. Value of the $\mathcal{S}$ factor for various configurations of Fabry-Perot cavity experiments

Configuration	Sensitivity Factor $\mathcal{S}$
Fig . 5 E_z	$\frac{\mu_r}{2}$
Fig . 5 H_z	$\frac{1}{2\epsilon_r}$
Fig . 5 Circular Polarization	$\frac{1}{4}\left(\frac{1}{\epsilon_r}+\mu_r\right)$
Fig . 6 (a)	$\frac{1}{2}\left(\frac{1}{\epsilon_r}-\mu_r\right)$
Fig . 6 (b)	$\frac{1}{4}\left(\frac{1}{\epsilon_r}-\mu_r\right)$

sensitive factor, $\mathcal{S}$, is the same as the vacuum cavity, while for the circularly polarized case, the sensitivity is close to that of the single vacuum cavity resonator. In contrast the same experiment with H_z polarization has reduced sensitivity of the order of the permittivity of the material. The sensitivity of the two Dual-Mode resonators gives the possibility of realizing a similar sensitivity to dual cavity experiments, but within the same cavity. The configuration should have a large degree of common mode rejection, and will be much more insensitive to external effects like temperature, vibration etc. and other systematics, and may be worth pursuing for these reasons. Note that Müller has recently completed an analysis of conventional cavity configurations in the electron (due to dispersion changes) and photon sector [32]. In our analysis we have only considered the photon sector and we have proposed some new unconventional configurations. It may be interesting to analyze these configurations in the electron sector. In the next subsection we consider similar configurations for Whispering Gallery (WG) modes.

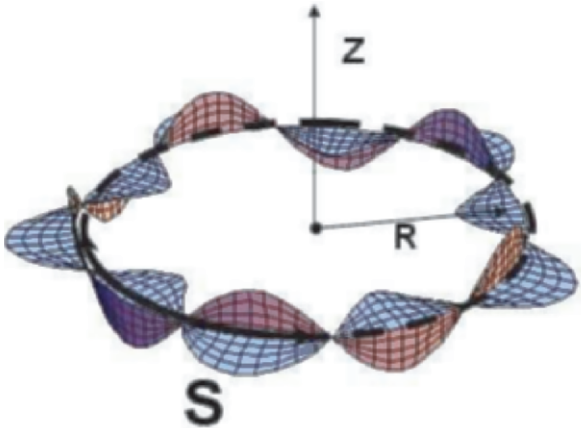


Fig. 7. A visual representation of the electric and magnetic fields of a "pure" Whispering Gallery mode propagating in the ϕ direction with a radius of $r = R$

4.2 Whispering Gallery Mode Resonators

In this subsection we consider "pure" WG modes, with the electric and magnetic fields propagating around with cylindrical symmetry at a discontinuity, with the direction of the Poynting vector ($\mathbf{E} \times \mathbf{B}$) as shown in Fig. 7. Thus, it is natural to analyse such modes in cylindrical coordinates $\{r, \phi, z\}$.

For an actual WG mode the wave is reflected off an electromagnetic discontinuity, and the fields mainly lie within the radius of the discontinuity and a smaller inner caustic [11]. However, by taking the limit as the azimuthal mode number m tends to infinity, the inner caustic converges to the radius of the discontinuity and the fields are reduced to a Dirac delta function. There are two possible polarizations, WGE with dominant H_z and E_r fields and WGH with dominant E_z and H_r. For "pure" WG modes, WGE have non-zero electric and magnetic filling factors of $Pe_r = 1$ and $Pm_z = 1$, and WGH have electric and magnetic filling factors of $Pe_z = 1$ and $Pm_r = 1$, in cylindrical coordinates. The electric and magnetic filling factors may be converted from cylindrical to cartesian symmetry by (the z component of the filling factor need not be transformed):

$$Pe_x = Pe_y = \frac{Pe_r + Pe_\phi}{2} \quad : \quad Pm_x = Pm_y = \frac{Pm_r + Pm_\phi}{2} \tag{47}$$

We can now do a similar analysis to Subsect. 4.1 for configurations shown in Figs. 8 and 9 for the WG case, with the computed sensitivities listed in Table 5. In vacuum the $\mathcal{S}$ factor is half that of the FP cavities in Subsect. 4.1, and the Dual-Mode resonator is insensitive. However, in a low loss dielectric the $\mathcal{S}$ factor approaches the same value for WGE modes as the FP cavity experiments, but the WGH modes remain about a factor of two less sensitive. The value of the

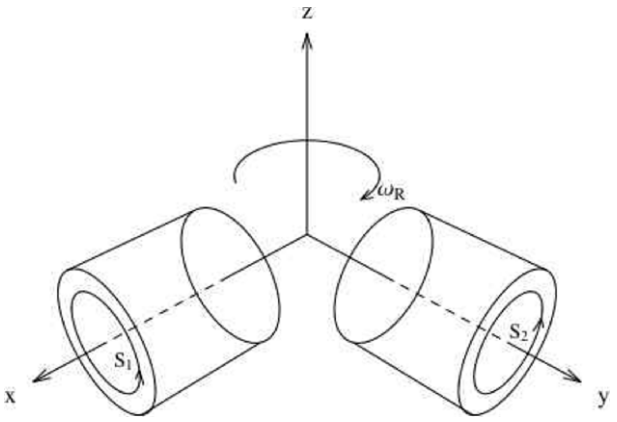


Fig. 8. Rotating Lorentz invariance test using two WG mode cavities

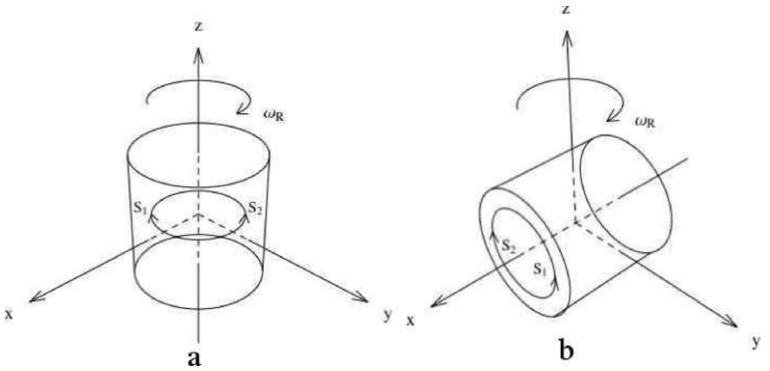


Fig. 9. Rotating Lorentz invariance test using two WG modes excited in one cavity (Dual-Mode resonator)

Table 5. Value of the $\mathcal{S}$ factor for various configurations of WG mode resonator cavity experiments

Configuration	Sensitivity Factor $\mathcal{S}$
Fig . 8 WGH	$-\frac{1}{2\epsilon_r}+\frac{\mu_r}{4}$
Fig . 8 WGE	$\frac{1}{4\epsilon_r}-\frac{\mu_r}{2}$
Fig . 9 (a)	0
Fig . 9 (b)	$\frac{3}{8}\left(\frac{1}{\epsilon_r}-\mu_r\right)$

$\mathcal{S}$ factor for the Dual-Mode resonator is the mean value of the WGE and WGH modes.

We have shown that similar sensitivities can be achieved with FP and WG cavity resonators. At UWA we have developed an experiment that uses low loss sapphire crystals, which exhibit a small uniaxial dielectric anistropy. The calculations of the sensitivity are presented Sect. 6.

5 Applying the RMS to Whispering Gallery Mode Resonator Experiments

In this section we restrict ourselves to analysis of whispering gallery mode resonator experiments, as the analysis has been well described for Fabry-Perot resonators previously [28]. For the whispering gallery mode experiment as shown in Fig. 8, the variable of integration around the path of the resonator is naturally chosen as the azimuthal angle, ϕ_j, relative to the cylindrical co-ordinates of each resonator. Thus, from (3) a frequency shift due to a putative Lorentz violation in the RMS framework is given by,

$$\frac{\Delta\nu_0}{\nu_0} = \frac{P_{MM}}{2\pi c^2}\left[\oint\left(\mathbf{v}.\hat{\mathbf{I}}_{\mathbf{a}}(\phi_{\mathbf{a}})\right)^2 d\phi_a - \oint\left(\mathbf{v}.\hat{\mathbf{I}}_{\mathbf{b}}(\phi_{\mathbf{b}})\right)^2 d\phi_b\right] \tag{48}$$

The dominant components of the velocity vector $\mathbf{v}$ were already calculated in Sect. 2.1, so to complete the calculation the time dependence of $\hat{\mathbf{I}}_a$ and $\hat{\mathbf{I}}_b$ must be calculated with respect to the MM-Earth frame. This of course depends on the sidereal and semi-sidereal frequencies, as well as the rotation frequency of the experiment. To start the calculation we define the time, $t = 0$ when the axis of the two WG resonators are aligned as shown in Fig. 8 (i.e. the resonators align with the laboratory frame). Then from this time we assume the resonator is rotated in a anti-clockwise direction of frequency ω_s, so the angle of rotation is $\gamma = \omega_s(t - t_s)$. Also, the longitudinal angle of the experiment is λ, which is dependent on the sidereal frequency and given by $\lambda = \omega_\oplus(t - t_l)$. Then we define the resonator with its cylinder axis in the y direction as resonator a, and the resonator with its cylinder axis in the x direction as resonator b. We also assume the WG modes are oscillating in a clockwise direction. In actual fact the calculation has been verified to be independent of the WG mode direction, and in most experiments is usually a standing wave (depending on the excitation) [33]. Thus in the laboratory frame at $t = 0$ the unit vectors in the direction of the Poynting vector are;

$$\mathbf{I}_a(\phi_a) = \begin{pmatrix} -\sin\phi_a \\ 0 \\ \cos\phi_a \end{pmatrix} \quad \mathbf{I}_b(\phi_b) = \begin{pmatrix} 0 \\ \cos\phi_b \\ \sin\phi_b \end{pmatrix} \tag{49}$$

Now if we transform from the resonator to the laboratory, then to the MM-Earth frame the unit vectors become.

$$\mathbf{I}_{Earth:a} = \begin{pmatrix} -\sin\phi_a(\cos\lambda\cos\chi\cos\gamma - \sin\gamma\sin\lambda) + \cos\phi_a\cos\lambda\sin\chi \\ -\sin\phi_a(\cos\lambda\sin\gamma + \cos\gamma\cos\chi\sin\lambda) + \cos\phi_a\sin\lambda\sin\chi \\ \cos\chi\cos\phi_a + \sin\chi\cos\gamma\sin\phi_a \end{pmatrix} \tag{50}$$

$$\mathbf{I}_{Earth:b} = \begin{pmatrix} -\sin\phi_b(\cos\lambda\cos\chi\sin\gamma + \cos\gamma\sin\lambda) + \cos\phi_b\cos\lambda\sin\chi \\ \sin\phi_b(\cos\lambda\cos\gamma - \sin\gamma\cos\chi\sin\lambda) + \cos\phi_b\sin\lambda\sin\chi \\ \cos\chi\cos\phi_b + \sin\chi\sin\gamma\sin\phi_b \end{pmatrix} \tag{51}$$

Here as in the previous sections χ is the angle from the north pole (co-latitude).

Table 6. Dominant coefficients in the RMS, using a short data set approximation calculated from (48)

ω_i	$10^7 Cu_{\omega_i}/P_{MM}$
$2\omega_R$	$-3.904 + 3.904\cos(\chi)^2 + 0.098\sin(\chi)^2$ $-\sin(\Phi_0)\left(-0.607 + 0.607\cos(\chi)^2\right)$ $-\cos(\Phi_0)\left(-0.120 + 0.120\cos(\chi)^2 - 0.055\sin(\chi)^2\right)$
$2\omega_R + \omega_\oplus$	$-0.876\sin(\chi) + 0.876\cos(\chi)\sin(\chi)$ $+\sin(\Phi_0)\left(0.068\sin(\chi) - 0.068\cos(\chi)\sin(\chi)\right)$ $+\cos(\Phi_0)\left(-0.232\sin(\chi) + 0.232\cos(\chi)\sin(\chi)\right)$
$2\omega_R - \omega_\oplus$	$0.876\sin(\chi) + 0.876\cos(\chi)\sin(\chi)$ $+\sin(\Phi_0)\left(-0.068\sin(\chi) - 0.068\cos(\chi)\sin(\chi)\right)$ $+\cos(\Phi_0)\left(0.232\sin(\chi) + 0.232\cos(\chi)\sin(\chi)\right)$
$2\omega_R + 2\omega_\oplus$	$1.952 - 3.904\cos(\chi) + 1.952\cos(\chi)^2$ $+\cos(\Phi_0)\left(-0.060 + 0.120\cos(\chi) - 0.060\cos(\chi)^2\right)$ $+\sin(\Phi_0)\left(-0.303 + 0.607\cos(\chi) - 0.303\cos(\chi)^2\right)$
$2\omega_R - 2\omega_\oplus$	$1.952 + 3.904\cos(\chi) + 1.952\cos(\chi)^2$ $+\cos(\Phi_0)\left(-0.060 - 0.120\cos(\chi) - 0.060\cos(\chi)^2\right)$ $+\sin(\Phi_0)\left(-0.303 - 0.607\cos(\chi) - 0.303\cos(\chi)^2\right)$

The next step is to substitute (50), (51) and (9) into (48). However, to be consistent with the SME analysis the phase should be calculated with respect to the vernal equinox, so that $\lambda_0 = \Phi_0 + \pi$ is substituted into (9) before we substitute it into (48) to calculate the frequency shift. Also, because we defined the rotation to be clockwise in the SME, to be consistent we define $\gamma_R = \omega_R(t - t_s)$ where $\omega_R = -\omega_s$. In this case the frequency components, which experience a frequency shift are given in Table 6. From the results of the calculation we note that perturbations due to Lorentz violations occur at the same frequencies as the SME (see Subsect. 7.2). Fortunately, it is not necessary to consider perturbations at exactly twice the spin frequency, $2\omega_s$, that are primarily due to the larger systematic effects associated with the rotation, as we only need to put a limit on one parameter. Also, the cosine components (Cu_{ω_i}) with respect to the CMB are the most sensitive, so we need not consider the sine components.

5.1 Phase with Respect to the CMB

To extract the P_{MM} term from our data we must first determine the phase of our experiment with respect to the CMB. Thus, in similar way to the reasoning for the SME (see Subsect. 3.2) we require δ_R, the difference between the experiment's

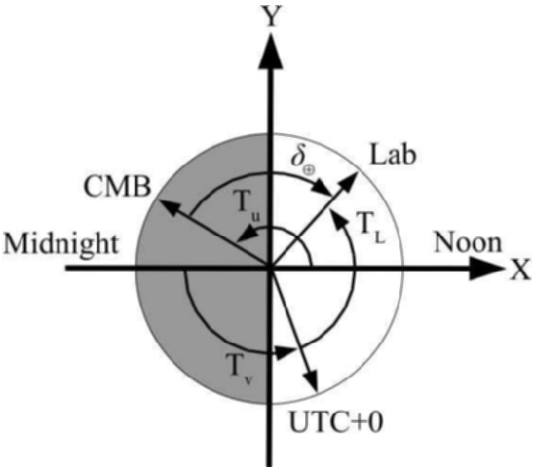


Fig. 10. Diagram showing meridians and angles used to determine the phase of the experiment with respect to the CMB

rotation time and the time since the vernal equinox, and $\delta_\oplus$, the difference between the sidereal time and the time since the vernal equinox.

As was the case for the SME, $\delta_R = 0$ since the axis of symmetry of the first resonator, a, is aligned with the local y-axis at $T_R = 0$. However, $\delta_\oplus$ will be different since in the case of the RMS it is measured with respect to the CMB (or MM-Earth frame), not the SCCEF. The CMB is oriented at 11.2 h right ascension, 6.4 degrees declination relative to the equatorial plane. Let us define T_v to be the time in seconds after midnight UTC+0, at which the vernal equinox has occurred in the J2000.0 frame [7]. T_u is the direction of the CMB (11.2h). For convenience we also define our longitude T_L in terms of sidereal seconds from midnight (in the case of our laboratory $T_l = 115.826° \times \frac{23\text{ h }56\text{ min}}{360°} = 27721$ s). As shown in Fig. 10, we now have an expression for $\delta_\oplus$.

$$\delta_\oplus = T_L + T_v - \left(T_u + \frac{23\text{ h }56\text{ min}}{2}\right) \tag{52}$$

Hence we are able to determine the phase of the experiment orientation relative to the CMB.

6 The University of Western Australia Rotating Experiment

Our experiment consists of two cylindrical sapphire resonators of 3 cm diameter and height supported by spindles at either end within superconducting niobium cavities [34], which are oriented with their cylindrical axes orthogonal to each other in the horizontal plane (see Fig. 11). Whispering gallery modes [35] are

Fig. 11. The two resonators are positioned orthogonal to each other in the mounting structure. One of the sapphires can be seen mounted inside the superconducting niobium cavity. The spindles are firmly held in each lid by sprung brass bushes

excited close to 10 GHz, with a difference frequency of 226 kHz. The frequencies are stabilized using a Pound locking scheme, and amplitude variations are suppressed using an additional control circuit. A detailed description of the cryogenic oscillators can be found in [36, 37], and a schematic of the experimental setup is shown in Fig. 12. The resonators are mounted in a common copper block, which provides common mode rejection of temperature fluctuations due to high thermal conductivity at cryogenic temperatures. The structure is in turn mounted inside two successive stainless steel vacuum cylinders from a copper post, which provides the thermal connection between the cavities and the liquid helium bath. A stainless steel section within the copper post provides thermal filtering of bath temperature fluctuations. A foil heater and carbon-glass

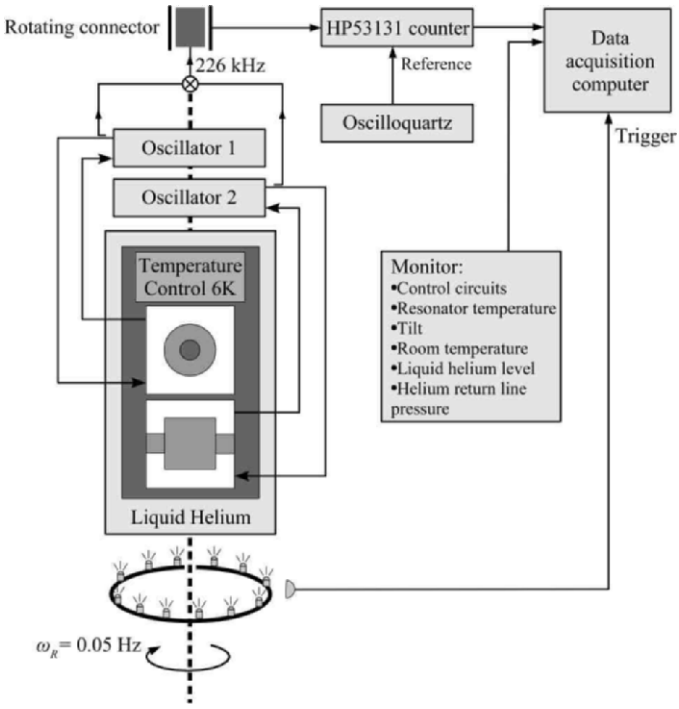


Fig. 12. Schematic of the experimental setup

temperature sensor attached to the copper post controls the temperature set point to 6 K with mK stability. Two stages of vacuum isolation are used to avoid contamination of the sapphire resonators from the microwave and temperature control devices located in the cryogenic environment.

A schematic of the rotation system is shown in Fig. 13. A cryogenic dewar containing the resonators, along with the room temperature oscillator circuits and control electronics, is suspended within a ring bearing. A multiple "V" shaped suspension made from loops of elastic shock cord avoids high Q-factor pendulum modes by ensuring that the cord has to stretch and shrink (providing damping losses) for horizontal motion as well as vertical. The rotation system is driven by a microprocessor controlled stepper motor. A commercial 18 conductor slip ring connector, with a hollow through bore, transfers power and various signals to and from the rotating experiment. A mercury based rotating coaxial connector transmits the difference frequency to a stationary frequency counter referenced to an Oscilloquartz oscillator. The data acquisition system logs the difference frequency as a function of orientation, as well as monitoring systematic effects including the temperature of the resonators, liquid helium bath level, ambient room temperature, oscillator control signals, tilt, and helium return line pressure.

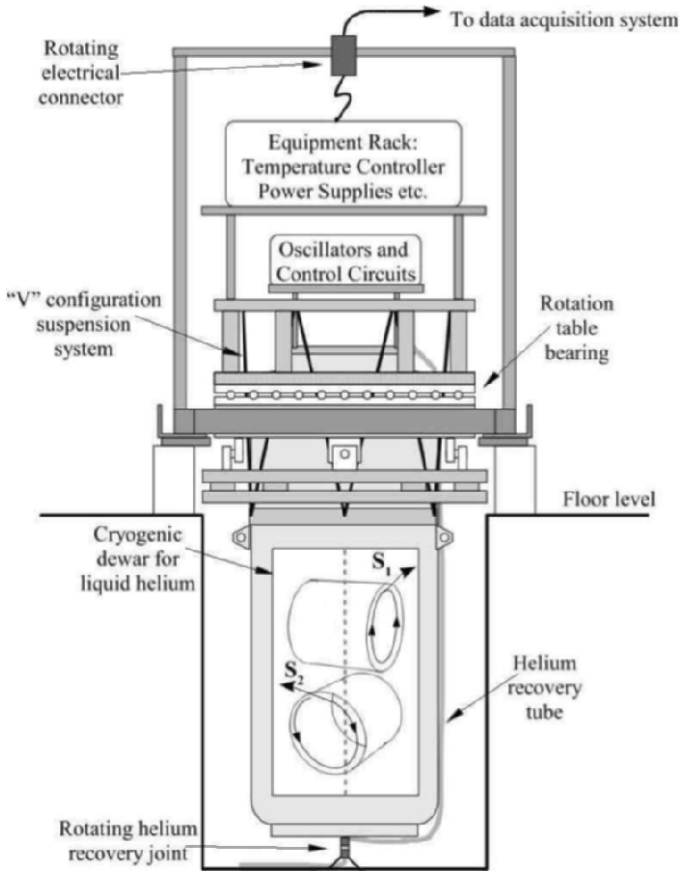


Fig. 13. Schematic of the cryogenic dewar, mounted in the rotation table. Inside the dewar a schematic of the two orthogonally orientated resonators is shown, along with the Poynting vectors of propagation S_1 and S_2

Inside the sapphire crystals standing waves are set up with the dominant electric and magnetic fields in the axial and radial directions respectively, corresponding to a propagation (Poynting) vector around the circumference. The observable of the experiment is the difference frequency, and to test for Lorentz violations the perturbation of the observable with respect to an alternative test theory must be derived. The mode which we excite is a Whispering Gallery mode we have a choice of $\mathrm{WGE}_{m,n,p}$ or $\mathrm{WGH}_{m,n,p}$ modes, the first subscript, m, gives the azimuthal mode number, while n and p give the number of zero crossings in the radial and z-direction respectively. Typically the so called fundamental mode families $\mathrm{WGE}_{m,0,0}$ or $\mathrm{WGH}_{m,0,0}$ as they have the highest Q-factors. To calculate the sensitivity in the RMS we use the technique presented in Sect. 5, while in the following subsection we numerically compute the sensitivity in the SME.

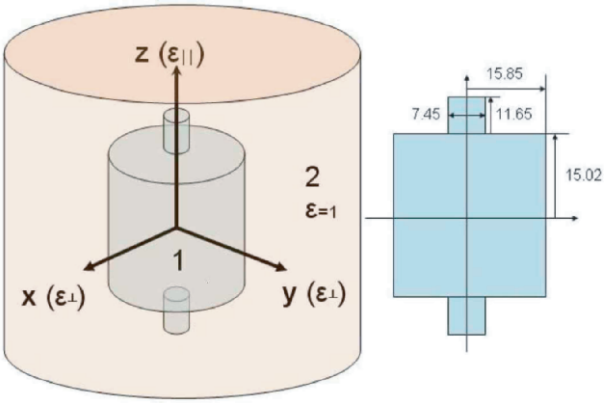


Fig. 14. Schematic of a cylindrical sapphire crystal resonator, with dimensions in mm. The crystal exhibits uniaxial anisotropy with the axis of symmetry (c-axis) aligned along the cylindrical z axis. The permittivity along the c-axis given by $\epsilon_{||}$. Perpendicular to the c-axis in the x, y or r, ϕ plane, the permittivity is given by $\epsilon_{\perp}$. The two regions shown are 1. the crystal and 2. the cavity

6.1 Sensitivity in the SME

In this subsection we calculate the sensitivity of the fundamental WG mode families, $\mathrm{WGE}_{m,0,0}$ and $\mathrm{WGH}_{m,0,0}$ to putative Lorentz violation in the SME, and compare it with the "pure" WG approximation given in Fig. 8. For a proper analysis of the sapphire loaded cavity resonators two regions of space need to be taken into account: the anisotropic crystal and the cavity free space surrounding it (see Fig. 14). The latter has a relative permittivity of 1, while both have relative permeability of 1 in all directions. The calculations proceed by splitting up V into V_1 (the crystal) and V_2 (the freespace), so we may sum the components of the $\mathcal{M}$ matricies over the two volumes (see (31) and (32)). The resonator operates close to liquid helium temperatures (6 Kelvin), where the permittivity of sapphire is, $\epsilon_{\perp} = 9.272$ and $\epsilon_{||} = 11.349$. To determine the sensitivity, we need to just calculate the experiments $\mathcal{S}$ factor in a similar way to the calculation for the "pure" WG modes in Subsect. 4.2. In this case the electric and magnetic filling factors must be calculated using a numeric technique such as finite element analysis, method of lines or separation of variables [11]. In this work we have chosen to use method of lines developed at IRCOM at the University of Limoges [39] . The calculated electric and magnetic field densities for the chosen mode ($WGH_{8,0,0}$) of operation at 10 GHz is shown in Fig. 15, and the $\mathcal{S}$ factor is calculated to be 0.19575.

The actual WG modes have all field components in both regions of the crystal. This modifies the sensitivity slightly, but approaches the limit of the "pure" WG mode as $m \to \infty$. The magnitude of the $\mathcal{S}$ factor for the fundamental WGE and WGH modes at X-Band (8GHz-12GHz) are plotted in Fig. 16. The WGH modes seem converge nicely toward the predicted "pure" WGH mode sensitivity,

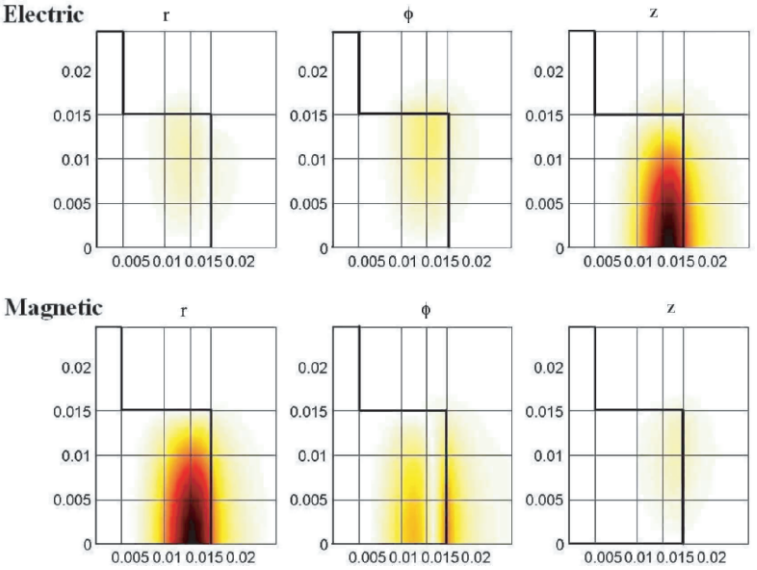


Fig. 15. The magnetic and electric field density plots shown in the top right-hand quadrant of an axial slice through the sapphire crystal and cavity for the $WGH_{8,0,0}$ mode. Note the dominant fields are E_z and H_r consistent with a pure whispering gallery mode approximation

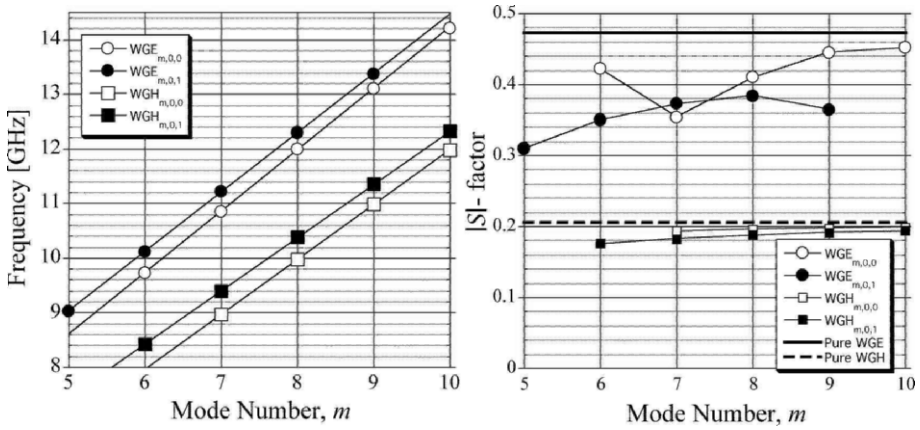


Fig. 16. The frequency and $|\mathcal{S}|$ factor as a function of mode number, m for the two lowest frequency WGE and WGH mode families. Note the general convergence to the "pure" WG value as m increases

while the WGE modes have a dip in sensitivity. This can be explained by an intersection with another mode of the same m number, resulting in a spurious mode interaction [38] . This does not occur in WGH modes since they are the lowest frequency modes for the mode number m (refer to Fig. 2 of [38]). It is

important to note that about a factor of two in sensitivity can be gained if we use a WGE mode rather than a WGH mode. However because we are using a 3 cm crystal rather than a 5 cm crystal, the Q-factor of WGE modes are degraded due to radiation and wall losses. In the future we can markedly improve the sensitivity by employing the typical 5 cm cavities that operate in WGE modes, as were used in the non-rotating experiments [11].

7 Data Analysis and Interpretation of Results

Figure 17 shows typical fractional frequency instability of the 226 kHz difference with respect to 10 GHz, and compares the instability when rotating and stationary. A minimum of 1.6×10^{-14} is recorded at 40 s. Rotation induced systematic effects degrade the stability up to 18s due to signals at the rotation frequency of 0.056 Hz and its harmonics. We have determined that tilt variations dominate the systematic by measuring the magnitude of the fractional frequency dependence on tilt and the variation in tilt at twice the rotation frequency, $2\omega_R(0.11\text{ Hz})$, as the experiment rotates. We minimize the effect of tilt by manually setting the rotation bearing until our tilt sensor reads a minimum at $2\omega_R$. The latter data sets were up to an order of magnitude reduced in amplitude as we became more experienced at this process. The remaining systematic signal is due to the residual tilt variations, which could be further annulled with an automatic tilt control system. It is still possible to be sensitive to Lorentz violations in the presence of these systematics by measuring the sidereal, $\omega_\oplus$, and semi-sidereal, $2\omega_\oplus$, sidebands about $2\omega_R$, as was done in [8]. The amplitude and phase of a Lorentz violating signal is determined by fitting the parameters of (53) to the data, with the phase of the fit adjusted according to the test theory used.

$$\frac{\Delta\nu_0}{\nu_0} = A + Bt + \sum_i C_{\omega_i}\cos(\omega_i t + \varphi_i) + S_{\omega_i}\sin(\omega_i t + \varphi_i) \tag{53}$$

Here ν_0 is the average unperturbed frequency of the two sapphire resonators, and $\Delta\nu_0$ is the perturbation of the 226 kHz difference frequency, A and B determine the frequency offset and drift, and C_{ω_i} and S_{ω_i} are the amplitudes of a cosine and sine at frequency ω_i respectively. In the final analysis we fit 5 frequencies to the data, $\omega_i = (2\omega_R, 2\omega_R \pm \omega_\oplus, 2\omega_R \pm 2\omega_\oplus)$, as well as the frequency offset and drift. The correlation coefficients between the fitted parameters are all between 10^{-2} to 10^{-5}. Since the residuals exhibit a significantly non-white behavior, the optimal regression method is weighted least squares (WLS) [13]. WLS involves pre-multiplying both the experimental data and the model matrix by a whitening matrix determined by the noise type of the residuals of an ordinary least squares analysis.

We have acquired 5 sets of data over a period of 3 months beginning December 2004, totaling 18 days. The length of the sets (in days) and size of the systematic are $(3.6, 2.3 \times 10^{-14})$, $(2.4, 2.1 \times 10^{-14})$, $(1.9, 2.6 \times 10^{-14})$, $(4.7, 1.4 \times 10^{-15})$, and $(6.1, 8.8 \times 10^{-15})$ respectively. We have observed leakage of the systematic into

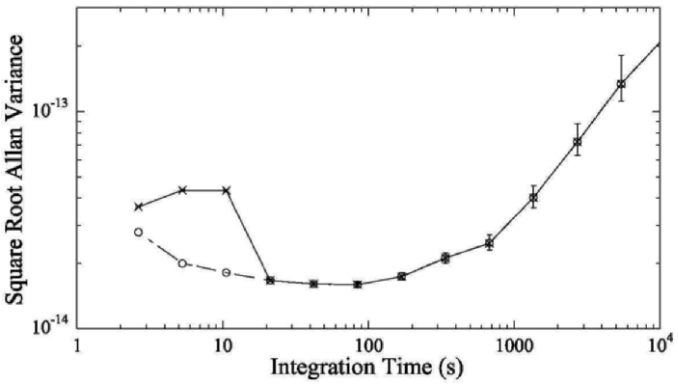


Fig. 17. Square Root Allan Variance fractional frequency instability measurement of the difference frequency when rotating (*crosses*) and stationary (*circles*). The hump at short integration times is due to systematic effects associated with the rotation of the experiment, with a period of 18 seconds. Above 18 seconds the instability is the same as when the experiment is stationary

the neighboring side bands due to aliasing when the data set is not long enough or the systematic is too large. Figure 18 shows the total amplitude resulting from a WLS fit to 2 of the data sets over a range of frequencies about $2\omega_R$. It is evident that the systematic of data set 1 at $2\omega_R$ is affecting the fitted amplitude of the sidereal sidebands $2\omega_R \pm \omega_\oplus$ due to its relatively short length and large systematics. By analyzing all five data sets simultaneously using WLS

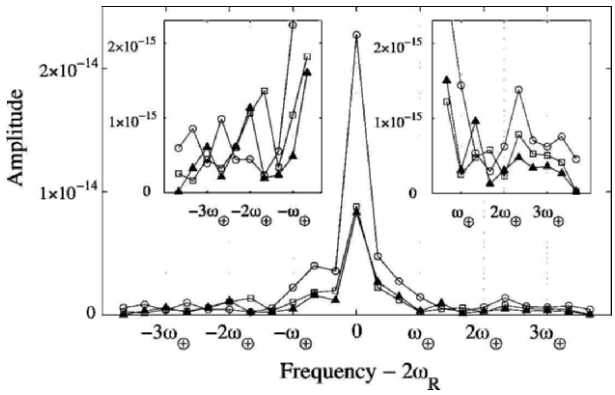


Fig. 18. Spectrum of amplitudes $\sqrt{C_{\omega_i}^2 + S_{\omega_i}^2}$ calculated using WLS, showing systematic leakage about $2\omega_R$ for 2 data sets, data set 1 (3.6 days, *circles*), data set 5 (6.1 days, *squares*) and the combined data (18 days spanning 3 months, *solid triangles*). Here $\omega_\oplus$ is the sidereal frequency (11.6 μHz). By comparing a variety of data sets we have seen that leakage is reduced in longer data sets with lower systematics. The insets show the typical amplitude away from the systematic, which have statistical uncertainties of order 10^{-16}

the effective length of the data is increased, reducing the width of the systematic sufficiently as to not contribute significantly to the sidereal and semi-sidereal sidebands.

7.1 Standard Model Extension Framework

In the photon sector of the SME 10 independent components of $\tilde{\kappa}_{e+}$ and $\tilde{\kappa}_{o-}$ have been constrained by astronomical measurements to $<2\times10^{-32}$ [7,23]. Seven components of $\tilde{\kappa}_{e-}$ and $\tilde{\kappa}_{o+}$ have been constrained in optical and microwave cavity experiments [10, 13] at the 10^{-15} and 10^{-11} level respectively, while the scalar $\tilde{\kappa}_{tr}$ component recently had an upper limit set of $<10^{-4}$ [24]. The remaining $\tilde{\kappa}_{e-}^{ZZ}$ component could not be previously constrained in non-rotating experiments [10, 13]. In contrast, our rotating experiment is sensitive to $\tilde{\kappa}_{e-}^{ZZ}$. However, it appears only at $2\omega_R$, which is dominated by systematic effects. By using the formulas derived in Table 3 for short data sets and the $\mathcal{S}$ factor for the $WGH_{8,0,0}$ mode in Fig. 16, the resulting numerical relation between the parameters of the SME and the C_{ω_i} and S_{ω_i} coefficients were calculated and are given in Table 7.

From our combined analysis of all data sets, and using the relation to $\tilde{\kappa}_{e-}^{ZZ}$ given in Table 7, we determine a value for $\tilde{\kappa}_{e-}^{ZZ}$ of $4.1(0.5)\times10^{-15}$. However, since we do not know if the systematic has canceled a Lorentz violating signal at $2\omega_R$, we cannot reasonably claim this as an upper limit. Since we have five individual data sets, a limit can be set by treating the $C_{2\omega_R}$ coefficient as a statistic. The phase of the systematic depends on the initial experimental conditions, and is random across the data sets. Thus, we have five values of $C_{2\omega_R}$, ($\{-4.2, 11.4, 21.4, 1.3, -8.1\}$ in 10^{-15}), two are negative coefficients and three are positive. If we take the mean of these coefficients, the systematic signal will cancel if the phase is random, but the possible Lorentz violating signal will not, since the phase is constant. Thus a limit can be set by taking the mean and standard deviation of the five coefficient of $C_{2\omega_R}$. This gives a more conservative bound of $2.1(5.7)\times10^{-14}$, which includes zero. Our experiment is also sensitive to all other seven components of $\tilde{\kappa}_{e-}$ and $\tilde{\kappa}_{o+}$ (see Table 7) and improves present limits by up to a factor of 7, as shown in Table 8.

7.2 Robertson, Mansouri, Sexl Framework

From (48), the dominant coefficients are calculated to be only due to the cosine terms with respect to the CMB right ascension, Cu_{ω_i}, and the theory predicts no perturbations in the quadrature term. Since our experiment is rotating clock wise we can substitute $2\omega_s = -2\omega_R$, and once we perform the integral and substitute all the numeric values. Following this method we calculate the coefficients as shown in Table 9.

The same five data sets were then re-analysed in the correct quadrature with respect to the CMB, with the results listed with the coefficients in Table 9. The measured and statistical uncertainty of P_{MM} is determined to be

Table 7. Coefficients C_{ω_i} and S_{ω_i} in (1) for the five frequencies of interest and their relation to the components of the SME parameters $\tilde{\kappa}_{e-}$ and $\tilde{\kappa}_{o+}$, derived using a short data set approximation including terms up to first order in orbital velocity, where Φ_0 is the phase of the orbit since the vernal equinox. Note that for short data sets the upper and lower sidereal sidebands are redundant, which reduces the number of independent measurements to 5. To lift the redundancy, more than a year of data is required so annual offsets may be de-correlated from the twice rotational and sidereal sidebands listed

ω_i	C_{ω_i}	S_{ω_i}
$2\omega_R$	$0.21\tilde{\kappa}_{e-}^{ZZ}$	-
$2\omega_R + \omega_\oplus$	$2.5 \times 10^{-5} \sin\Phi_0 \tilde{\kappa}_{o+}^{XY}$ $-1.0 \times 10^{-5} \cos\Phi_0 \tilde{\kappa}_{o+}^{YZ}$ $-0.27\ \tilde{\kappa}_{e-}^{XZ}$	$-\cos\Phi_0 \left[2.3 \times 10^{-5} \tilde{\kappa}_{o+}^{XY} - 1.0 \times 10^{-5} \tilde{\kappa}_{o+}^{XZ}\right]$ $-0.27\tilde{\kappa}_{e-}^{YZ}$
$2\omega_R + 2\omega_\oplus$	$-2.1 \times 10^{-5} \cos\Phi_0 \tilde{\kappa}_{o+}^{XZ}$ $+2.3 \times 10^{-5} \sin\Phi_0 \tilde{\kappa}_{o+}^{YZ}$ $-0.11(\tilde{\kappa}_{e-}^{XX} - \tilde{\kappa}_{e-}^{YY})$	$-2.3 \times 10^{-5} \sin\Phi_0 \tilde{\kappa}_{o+}^{XZ}$ $-2.1 \times 10^{-5} \cos\Phi_0 \tilde{\kappa}_{o+}^{YZ} - 0.23\ \tilde{\kappa}_{e-}^{XY}$
$2\omega_R - \omega_\oplus$	$-0.31 C_{2\omega_R+\omega_\oplus}$	$0.31 S_{2\omega_R+2\omega_\oplus}$
$2\omega_R - 2\omega_\oplus$	$9.4 \times 10^{-2} C_{2\omega_R+2\omega_\oplus}$	$-9.4 \times 10^{-2} S_{2\omega_R+2\omega_\oplus}$

Table 8. Results for the SME Lorentz violation parameters, assuming no cancellation between the isotropy terms. $\tilde{\kappa}_{e-}$ (in 10^{-15}) and first order boost terms $\tilde{\kappa}_{o+}$ (in 10^{-11}) [12]

	$\tilde{\kappa}_{e-}^{XY}$	$\tilde{\kappa}_{e-}^{XZ}$	$\tilde{\kappa}_{e-}^{YZ}$	$(\tilde{\kappa}_{e-}^{XX} - \tilde{\kappa}_{e-}^{YY})$
this work	-0.63(0.43)	0.19(0.37)	-0.45(0.37)	-1.3(0.9)
from [13]	-5.7(2.3)	-3.2(1.3)	-0.5(1.3)	-3.2(4.6)
	$\tilde{\kappa}_{e-}^{ZZ}$	$\tilde{\kappa}_{o+}^{XY}$	$\tilde{\kappa}_{o+}^{XZ}$	$\tilde{\kappa}_{o+}^{YZ}$
this work	21(57)	0.20(0.21)	-0.91(0.46)	0.44(0.46)
from [13]	–	-1.8(1.5)	-1.4(2.3)	2.7(2.2)

$-0.9(2.0) \times 10^{-10}$, which represents a factor of 7.5 improvement over previous results $2.2(1.5) \times 10^{-9}$ [10].

8 Summary

Rotating resonator experiments are emerging as one of the most sensitive types of Local Lorentz Invariance tests in electrodynamics (also see other contributions within these proceedings). In this work we have analysed in detail such

Table 9. Dominant coefficients in the RMS, using a short data set approximation calculated from (48). The measured values of P_{MM} (in 10^{-10}) are shown together with the statistical uncertainties in the bracket. From this data the measured and statistical uncertainty of P_{MM} is determined to be $-0.9(2.0)\times10^{-10}$, which represents more than a factor of 7.5 improvement over previous results $2.2(1.5)\times10^{-9}$ [10]

ω_i	Cu_{ω_i}	P_{MM}
$2\omega_R+\omega_\oplus$	$[-1.13\times10^{-7}-3.01\times10^{-8}\cos\Phi_0$ $+8.83\times10^{-9}\sin\Phi_0]P_{MM}$	$-2.1(7.2)$
$2\omega_R-\omega_\oplus$	$[3.51\times10^{-8}+9.31\times10^{-9}\cos\Phi_0$ $-2.73\times10^{-9}\sin\Phi_0]P_{MM}$	$62.4(23.3)$
$2\omega_R+2\omega_\oplus$	$[4.56\times10^{-7}-1.39\times10^{-8}\cos\Phi_0$ $-7.08\times10^{-8}\sin\Phi_0]P_{MM}$	$-1.3(2.1)$
$2\omega_R-2\omega_\oplus$	$[4.37\times10^{-8}-1.34\times10^{-9}\cos\Phi_0$ $-6.78\times10^{-9}\sin\Phi_0]P_{MM}$	$-7.5(22.1)$

experiments to putative Lorentz violation in both the RMS and SME frameworks. In the RMS, rotating experiments only enhance the sensitivity to the Michelson-Morley parameter, P_{MM}, and are not sensitive to the Kennedy-Thorndike, P_{KT}, or Ives-Stilwell, P_{IS}, parameters. In the SME non-rotating resonator experiments in the laboratory test for four components of the $\tilde{\kappa}_{e-}$ tensor and three components of the $\tilde{\kappa}_{o+}$ tensor, with the scalar coefficient $\tilde{\kappa}_{tr}$ and $\tilde{\kappa}_{e-}^{ZZ}$ unmeasurable. Rotation in the SME enhances the sensitivity to the seven components and also allows the determination of the $\tilde{\kappa}_{e-}^{ZZ}$ component. We have shown that all resonator experiment exhibit the same relative frequency spectrum to the putative signal to within a multiplicative sensitivity factor, $\mathcal{S}$. This was utilized to compare the sensitivity of different FP and WG resonator configurations, leading to the proposal of some new dual-mode resonator experiments.

We applied the above analysis to our experiment at the University of Western Australia, which is based on rotating cryogenic sapphire whispering gallery mode microwave oscillators. In summary, we presented the first results of the experiment, which we set bounds on 7 components of the SME photon sector (Table 8) and P_{MM} (Table 9) of the RMS framework up to a factor of 7.5 more stringent than those obtained from previous experiments. We also set an upper limit $(2.1(5.7)\times10^{-14})$ on the previously unmeasured SME component $\tilde{\kappa}_{e-}^{ZZ}$. To further improve these results, tilt and environmental controls will be implemented to reduce systematic effects. To remove the assumption that the $\tilde{\kappa}_{o+}$ and $\tilde{\kappa}_{e-}$ do not cancel each other, data integration will continue for more than a year.

Acknowledgments

This work was funded by the Australian Research Council.

References

1. Will C.M., *Theory and Experiment in Gravitational Physics*, revised edition (Cambridge University Press, Cambridge, 1993).
2. Robertson H.P., Rev. Mod. Phys. **21**, 378 (1949).
3. Mansouri R., Sexl R.U., Gen. Rel. Grav. **8**, 497 (1977).
4. Lightman A.P., Lee D.L., Phys. Rev. **D 8**, 364 (1973).
5. Ni W.-T., Phys. Rev. Lett. **38**, 301 (1977).
6. Colladay D., Kostelecký V.A., Phys.Rev. **D 55**, 6760 (1997); Colladay D., Kostelecký V.A., Phys. Rev. **D58**, 116002 (1998)
7. Kostelecký V.A.,Mewes M., Phys. Rev. **D 66**, 056005, (2002).
8. Brillet A., Hall J., Phys. Rev. Lett. **42**, 549 (1979).
9. Wolf P. et al., Phys. Rev. Lett. **90**, 060402 (2003).
10. Müller H. et al., Phys. Rev. Lett. **91**, 020401 (2003).
11. Wolf P., et al., Gen. Rel. and Grav., **36**, 2351 (2004).
12. Lipa J.A. et al., Phys. Rev. Lett. **90**, 060403 (2003).
13. Wolf P, et al., Phys. Rev. **D 70**, 051902(R), (2004).
14. Tobar M.E. et al., Phys. Lett. A **300**, 33 (2002).
15. Stanwix P.L., Tobar, M.E., Wolf P. et al., Phys. Rev. Lett. **95**, 040404 (2005).
16. Antonini P. et al., Phys. Rev. A. **71**, 050101 (2005).
17. Herrmann S. et al., arXiv: physics 050809, accepted for publication in Phys. Rev. Lett., (2005).
18. Michelson A.A. and Morley E.W., Am. J. Sci. **34**, 333 (1887).
19. Kennedy R.J. and Thorndike E.M., Phys. Rev. **B 42**, 400 (1932).
20. Ives H.E. and Stilwell G.R., J. Opt. Soc. Am. **28**, 215 (1938).
21. Saathoff G., et al., Phys. Rev. Lett **91**, 190403 (2003).
22. Jackiw R. and Kostelecký V.A., Phys. Rev. Lett, **82** (1999).
23. Kostelecký V.A., and Mewes M., Phys. Rev. Lett. **87**, 251304 (2001).
24. Tobar M.E. et al., Phys. Rev. **D71**, 025004 (2005).
25. Bluhm R. et al., Phys. Rev. **D68**, 125008 (2003).
26. Bailey Q.G., and Kostelecký V.A., Phys. Rev. **D70**, 076006 (2004).
27. Krupka J., Derzakowski K., Abramowicz A., Tobar M.E. and Geyer R., IEEE Trans. on Microw. Theory Tech. **47,** 752–759, (1999).
28. H. Müller, S. Herrmann, C. Braxmaier, S. Schiller and A. Peters, Appl Phys. B (Laser Opt.) **77,** 719–773, (2004).
29. Hall J.L., Ye J., and Ma L.-S., Phys. Rev. **A62**, 013815 (2000).
30. Tobar M.E., Hartnett J.G., Phys. Rev. **D67**, 062001 (2003).
31. Braxmaier C., Pradl O., Müller H., Peters A., Mlynek J., Loriette V., Schiller S., Phys. Rev. **D64**, 042001 (2001).
32. Müller H., Phys. Rev. **D71**, 045004 (2003).
33. Tobar M.E. et al., IEEE Trans. on Ultrason., Ferroelect. Fr Contr., **47**, 421, (2000).
34. Giles A.J. et al., Physica B **165**, 145, (1990).
35. Tobar M.E., Mann A.G., IEEE Trans. Microw. Theory Tech. **39,** 2077, (1991).
36. Mann A.G., Topics in Appl. Phys. **79**, 37, (2001).
37. Hartnett J.G., Tobar M.E., Topics in Appl. Phys. **79**, 67, (2001).
38. Tobar M.E., Hartnett J.G., Ivanov E.N., Cros D., Blondy P., IEEE Trans. Instr. and Meas. **30,** 522, (2001).
39. Tobar M.E. et al., IEEE Trans. Ultrason., Ferroelect. Fr Contr., **52**, 17, (2005).

Recent Experimental Tests of Special Relativity

P. Wolf[1,2], S. Bize[1], M.E. Tobar[3], F. Chapelet[1], A. Clairon[1], A.N. Luiten[3], and G. Santarelli[1]

[1] BNM-SYRTE, Observatoire de Paris, 61 avenue de l'Observatoire, 75014 Paris, France
[2] Bureau International des Poids et Mesures, Pavillon de Breteuil, 92312 Sèvres Cedex, France
[3] University of Western Australia, School of Physics, Nedlands 6907 WA, Australia

Abstract. We review our recent Michelson-Morley (MM) and Kennedy-Thorndike (KT) experiment, which tests Lorentz invariance in the photon sector, and report first results of our ongoing atomic clock test of Lorentz invariance in the matter sector.

The MM-KT experiment compares a cryogenic microwave resonator to a hydrogen maser, and has set the most stringent limit on a number of parameters in alternative theories to special relativity. In the Robertson-Mansouri-Sexl (RMS) framework our experiment constrains $1/2 - \beta_{MS} + \delta_{MS} = (1.2 \pm 2.2) \times 10^{-9}$ and $\beta_{MS} - \alpha_{MS} - 1 = (1.6 \pm 3.0) \times 10^{-7}$, which is of the same order as the best results from other experiments for the former and represents a 70 fold improvement for the latter. In the photon sector of the general Lorentz violating standard model extension (SME), our experiment limits 4 components of the $\tilde{\kappa}_{e-}$ parameter to a few parts in 10^{-15} and the three components of $\tilde{\kappa}_{o+}$ to a few parts in 10^{-11}. This corresponds to an improvement by up to a factor 10 on best previous limits.

We also report first results of a test of Lorentz invariance in the SME matter sector, using Zeeman transitions in a laser cooled Cs atomic fountain clock. We describe the experiment together with the theoretical model and analysis. Recent experimental results are presented and analyzed, including statistical uncertainties and a brief discussion of systematic effects. Based on these results, we give a first estimate of components of the $\tilde{c}^p$ parameters of the SME matter sector. A full analysis of systematic effects is still in progress, and will be the subject of a future publication together with our final results. If confirmed, the present limits would correspond to first ever measurements of some $\tilde{c}^p$ components, and improvements by 11 and 14 orders of magnitude on others.

1 Introduction

One hundred years after Einstein's first paper [1] special relativity is still standing up to all experimental tests and verifications. Over the last century a large number of such tests have provided what is certainly one of the most solid experimental bases of any present fundamental theory of physics. As a consequence

P. Wolf et al.: *Recent Experimental Tests of Special Relativity*, Lect. Notes Phys. **702**, 451–478 (2006)
DOI 10.1007/3-540-34523-X_16 © Springer-Verlag Berlin Heidelberg 2006

special relativity is today underpinning all of present day physics, ranging from the standard model of particle physics (including nuclear and atomic physics) to general relativity and astronomy. That fact continues to push experimentalists to search for new experiments, or improve on previous ones, in order to uncover a possible violation of special relativity, as that would most certainly lead the way to a new conception of physics and of the universe surrounding us. Additional incentive for such tests comes from unification theories (e.g. string theories, loop quantum gravity), some of which [2–4] suggest a violation of special relativity at some, a priori unknown, level. Given the strong theoretical motivation for such theories, but the lack of experimental data that would allow a more rigorous selection among the candidate theories and the parameter space of each class of such theories, any experimental results that could aid the theoretical efforts are certainly welcome.

The fundamental hypothesis of special relativity is what Einstein termed the "principle of relativity" [1], or in more modern terms Local Lorentz Invariance (LLI) [5]. Loosely stated, LLI postulates that the outcome of any local test experiment is independent of the velocity of the freely falling apparatus. LLI can be viewed as a constituent part of the Einstein Equivalence Principle which is fundamental to general relativity and all metric theories of gravitation [5]. The experiments presented in this paper test some aspect of LLI, as characterized in Lorentz violating theoretical frameworks like the ones briefly described in Sect. 2.

We review and present two of our recent and ongoing experiments [7–9] that test different aspects of LLI, analyzing and describing their outcome in two theoretical frameworks, the kinematical test theory of Robertson, Mansouri and Sexl (RMS) [10, 11] and the Lorentz violating extension of the standard model (SME) [12]. These experiments, a Michelson-Morley and Kennedy-Thorndike test (Sect. 3), and an ongoing atomic clock test in the SME matter sector (Sect. 4), are among the most precise LLI tests at present.

The vast majority of modern experiments that test LLI rely essentially on the stability of atomic clocks and macroscopic resonators, therefore improvements in oscillator technology have gone hand in hand with improved tests of LLI. The experiments presented here are no exception. All of them employ clocks and resonators developed and used primarily for other purposes (national and international time scales, frequency calibration, etc.) but adapted for tests of LLI.

2 Theoretical Frameworks

Numerous test theories that allow the modeling and interpretation of experiments that test LLI have been developed. Kinematical frameworks [10, 11] postulate a simple parametrisation of the Lorentz transformations with experiments setting limits on the deviation of those parameters from their special relativistic values. A more fundamental approach is offered by theories that parametrise the coupling between gravitational and non-gravitational fields (TH$\epsilon\mu$ [5, 15, 16]

or χg [17] formalisms) which allow the comparison of experiments that test different aspects of the EEP. Finally, formalisms motivated by unification theories [3, 12, 18] have the advantage of opening the way to experimental investigations in the domain of the unification of gravity with the other fundamental forces of nature. In this work we restrict ourselves to two theoretical frameworks, the kinematical framework developed by Robertson, Mansouri and Sexl (RMS) and the more recent standard model extension (SME) of Kostelcký and co-workers.

By construction, kinematical frameworks do not allow for any dynamical effects on the measurement apparatus. This implies that in all inertial frames two clocks of different nature (e.g. based on different atomic species) run at the same relative rate, and two length standards made of different materials keep their relative lengths. Coordinates are defined by the clocks and length standards, and only the transformations between those coordinate systems are modified. In general this leads to observable effects on light propagation in moving frames but, by definition, to no observable effects on clocks and length standards. In particular, no attempt is made at explaining the underlying physics (e.g. modified Maxwell and/or Dirac equations) that could lead to Lorentz violating light propagation but leave e.g. atomic energy levels unchanged. On the other hand dynamical frameworks (e.g. the TH$\epsilon\mu$ formalism or the SME) in general use a modified general Lagrangian that leads to modified Maxwell and Dirac equations and hence to Lorentz violating light propagation and atomic properties, which is why they are considered more fundamental and more complete than the kinematical frameworks. Furthermore, as shown in [19], the SME is kept sufficiently general to, in fact, encompass the kinematical frameworks and some other dynamical frameworks (in particular the TH$\epsilon\mu$ formalism) as special cases, although there are no simple and direct relationships between the respective parameters.

2.1 The Robertson, Mansouri and Sexl Framework

Kinematical frameworks for the description of Lorentz violation have been pioneered by Robertson [10] and further refined by Mansouri and Sexl [11] and others. Fundamentally the different versions of these frameworks are equivalent, and relations between their parameters are readily obtained. As mentioned above these frameworks postulate generalized transformations between a preferred frame candidate $\Sigma(T, \mathbf{X})$ and a moving frame $S(t, \mathbf{x})$ where it is assumed that in both frames coordinates are realized by identical standards. The transformations of [11] (in differential form) for the case where the velocity of S as measured in Σ is along the positive X-axis, and assuming Einstein synchronization in S (in all of the following the choice of synchronization convention plays no role) are

$$dT = \frac{1}{a}\left(dt + \frac{vdx}{c^2}\right); dX = \frac{dx}{b} + \frac{v}{a}\left(dt + \frac{vdx}{c^2}\right); dY = \frac{dy}{d}; dZ = \frac{dz}{d} \qquad (1)$$

with c the velocity of light in vacuum in Σ, and $\mathbf{v}$ the velocity of S in Σ. In special relativity $\alpha_{\rm MS} = -1/2; \beta_{\rm MS} = 1/2; \delta_{\rm MS} = 0$ and (1) reduces to the usual Lorentz transformations. Generally, the best candidate for Σ is taken to be the frame of the cosmic microwave background (CMB) [20, 21] with the velocity of the solar system in that frame taken as $v_\odot \approx$377 km/s, decl. $\approx -6.4°$, $RA \approx$11.2h.

Michelson-Morley type experiments [22] determine the coefficient $P_{MM} = (1/2 - \beta_{\rm MS} + \delta_{\rm MS})$ of the direction dependent term. For many years the most stringent limit on that parameter was $|P_{MM}| \leq 5\times10^{-9}$ determined over 23 years ago in an outstanding experiment [23]. Our experiment [8] confirms that result with roughly equivalent uncertainty (2.2×10^{-9}). Recently an improvement to $|P_{MM}| \leq 1.5\times10^{-9}$ has been reported [25]. Kennedy-Thorndike experiments [24] measure the coefficient $P_{KT} = (\beta_{\rm MS} - \alpha_{\rm MS} - 1)$ of the velocity dependent term. The most stringent limit [27] on $|P_{KT}|$ has been recently improved from [28] by a factor 3 to $|P_{KT}| \leq 2.1 \times 10^{-5}$. Our experiment [8] improves this result by a factor of 70 to $|P_{KT}| \leq 3.0\times10^{-7}$. Finally Ives-Stilwell experiments [26] measure $\alpha_{\rm MS}$. The most stringent result comes from the recent experiment of [14] which improves by a factor 4 our 1997 results [6], limiting $|\alpha_{\rm MS} + 1/2|$ to $\leq 2.2\times10^{-7}$. The three types of experiments taken together then completely characterize any deviation from Lorentz invariance in this particular test theory, with present limits summarized in Table 1 (but note that Table 1 does not include new limits reported in these proceedings).

Table 1. Present limits on Lorentz violating parameters in the framework of [11], not including new limits reported in these proceedings

Reference	$\alpha_{\rm MS} + 1/2$	$1/2 - \beta_{\rm MS} + \delta_{\rm MS}$	$\beta_{\rm MS} - \alpha_{\rm MS} - 1$
Saathoff et al. 2003 [14]	$\leq 2.2 \times 10^{-7}$	–	–
Müller et al. 2003 [25]	–	$(2.2 \pm 1.5) \times 10^{-9}$	–
Braxmaier et al. 2002 [27]	–	–	$(1.9 \pm 2.1) \times 10^{-5}$
Wolf and Petit 1997 [6]	$\leq 8 \times 10^{-7}$	–	–
Wolf et al. 2003 [8]	–	$(1.2 \pm 2.2) \times 10^{-9}$	$(1.6 \pm 3.0) \times 10^{-7}$

2.2 The Standard Model Extension

The general Lorentz violating Standard Model Extension (SME) was developed relatively recently by Kostelecký and co-workers [12], motivated initially by possible Lorentz violating phenomenological effects of string theory [2]. It consists of a parametrised version of the standard model Lagrangian that includes all Lorentz violating terms that can be formed from known fields, and includes (in its most recent version [29]) gravity.

The fundamental theory of the SME as applied to electrodynamics is laid out in [19] and summarized below. We use that approach to model the MM and KT experiments in Sect. 3.2. For the discussion of the atomic clock experiment of Sect. 4 the SME matter sector is relevant. Its application to atomic physics, and in particular atomic clock experiments, is laid out in [30, 31] and summarized below.

Generally, the SME characterizes a potential Lorentz violation using a number of parameters that are all zero in standard (non Lorentz violating) physics. These parameters are frame dependent and consequently vary as a function of the coordinate system chosen to analyze a given experiment. In principle they may be constant and non-zero in any frame (e.g. the lab frame). However, any non-zero values are expected to arise from Planck-scale effects in the early Universe. Therefore they should be constant in a cosmological frame (e.g. the one defined by the CMB radiation) or any frame that moves with a constant velocity and shows no rotation with respect to the cosmological one. Consequently the conventionally chosen frame to analyze and compare experiments in the SME is a sun-centered, non-rotating frame as defined in [19]. The general procedure is to calculate the SME perturbation of the experimental observable in the lab frame (or cavity frame, or atom frame) and then to transform the lab frame SME parameters to the conventional sun-centered frame. This transformation will introduce a time variation of the frequency related to the movement of the lab with respect to the sun-centered frame (typically introducing time variations of sidereal and semi-sidereal periods for an Earth fixed experiment).

SME Photon Sector

The photon sector of the SME is described by a Lagrangian that takes the form

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}(k_{AF})^{\kappa}\epsilon_{\kappa\lambda\mu\nu}A^{\lambda}F^{\mu\nu} - \frac{1}{4}(k_F)_{\kappa\lambda\mu\nu}F^{\kappa\lambda}F^{\mu\nu} \tag{2}$$

where $F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu$. The first term is the usual Maxwell part while the second and third represent Lorentz violating contributions that depend on the parameters k_{AF} and k_F. For most analysis the k_{AF} parameter is set to 0 for theoretical reasons (c.f. [19]), which is also well supported experimentally. The remaining dimensionless tensor $(k_F)_{\kappa\lambda\mu\nu}$ has a total of 19 independent components that need to be determined by experiment. Retaining only this term leads to Maxwell equations that take the familiar form but with **D** and **H** fields defined by a general matrix equation

$$\begin{pmatrix} \mathbf{D} \\ \mathbf{H} \end{pmatrix} = \begin{pmatrix} \epsilon_0(\widetilde{\epsilon_r} + \kappa_{DE}) & \sqrt{\frac{\epsilon_0}{\mu_0}}\kappa_{DB} \\ \sqrt{\frac{\epsilon_0}{\mu_0}}\kappa_{HE} & \mu_0^{-1}(\widetilde{\mu_r^{-1}} + \kappa_{HB}) \end{pmatrix} \begin{pmatrix} \mathbf{E} \\ \mathbf{B} \end{pmatrix} \tag{3}$$

where the κ are 3×3 matrices whose components are particular combinations of the k_F tensor (c.f. equation (5) of [19]). If we suppose the medium of interest has general magnetic or dielectric properties, then $\widetilde{\epsilon_r}$ and $\widetilde{\mu_r}$ are also 3×3 matrices. In vacuum $\widetilde{\epsilon_r}$ and $\widetilde{\mu_r}$ are identity matrices. Equation (3) indicates a

useful analogy between the SME in vacuum and standard Maxwell equations in homogeneous anisotropic media.

For the analysis of different experiments it turns out to be useful to introduce further combinations of the κ matrices defined by:

$$\begin{aligned}
(\tilde{\kappa}_{e+})^{jk} &= \frac{1}{2}(\kappa_{DE} + \kappa_{HB})^{jk} \,, \\
(\tilde{\kappa}_{e-})^{jk} &= \frac{1}{2}(\kappa_{DE} - \kappa_{HB})^{jk} - \frac{1}{3}\delta^{jk}(\kappa_{DE})^{ll}, \\
(\tilde{\kappa}_{o+})^{jk} &= \frac{1}{2}(\kappa_{DB} + \kappa_{HE})^{jk} \,, \\
(\tilde{\kappa}_{o-})^{jk} &= \frac{1}{2}(\kappa_{DB} - \kappa_{HE})^{jk} \,, \\
\tilde{\kappa}_{tr} \quad &= \frac{1}{3}(\kappa_{DE})^{ll} \,. \qquad (4)
\end{aligned}$$

The first four of these equations define traceless 3×3 matrices, while the last defines a single coefficient. All $\tilde{\kappa}$ matrices are symmetric except $\tilde{\kappa}_{o+}$ which is antisymmetric. These characteristics leave a total of 19 independent coefficients of the $\tilde{\kappa}$.

In general experimental results are quoted and compared using the $\tilde{\kappa}$ parameters rather than the original k_F tensor components. The 10 independent components of the $\tilde{\kappa}_{e+}$ and $\tilde{\kappa}_{o-}$ tensors, have been determined to $\leq 2 \times 10^{-32}$ by astrophysical tests [19]. Of the 9 remaining independent components, 4 components of $\tilde{\kappa}_{e-}$ and the 3 components of $\tilde{\kappa}_{o+}$ have been bounded by the resonator experiments reported here and in [9, 25] to parts in 10^{15} and 10^{11} respectively, with our results improving by up to a factor 10 on the best previous ones (c.f. Table 7). The scalar $\tilde{\kappa}_{tr}$ has been bounded recently by our SME analysis [13] of the experiment of [14] to parts in 10^{-5}. In [13] we also propose several interferometer and resonator experiments that could improve the limit on $\tilde{\kappa}_{tr}$ to parts in 10^{11} and the limits on $\tilde{\kappa}_{o+}$ to parts in 10^{15}. Finally, the remaining component $\tilde{\kappa}_{e-}^{ZZ}$ is undetermined at present as it is not accessible to Earth fixed experiments. However, it should be accessible to experiments that are rotating in the laboratory, like the ones reported elsewhere in these proceedings, which should yield the first limits on that parameter and thereby complete the coverage of the parameter space in the SME photon sector. Present limits are summarized in Table 2 (not including new limits reported in these proceedings).

SME Matter Sector

In the matter sector, the SME modifies the Lagrangian of a spin 1/2 fermion [30, 40] via a number of parameterized Lorentz violating terms. When applied to atomic physics, this leads to a perturbation of the standard model Hamiltonian parametrised by 40 parameters for each fundamental particle (proton, neutron, electron), which in turn leads to a shift of the atomic energy levels and atomic transition frequencies (see [30, 31] for details). Quite generally, the energy level shifts can be expressed in the form

Table 2. Present limits on Lorentz violating parameters in the SME photon sector, not including new limits reported in these proceedings

Parameter	$\tilde{\kappa}_{e+}$	$\tilde{\kappa}_{o-}$	$\tilde{\kappa}_{e-}$	$(\tilde{\kappa}_{e-}^{ZZ})$	$\tilde{\kappa}_{o+}$	$\tilde{\kappa}_{tr}$
No. of components	5	5	4	1	3	1
Limits	10^{-32}	10^{-32}	10^{-15}	–	10^{-11}	10^{-5}
Reference	[19]	[19]	[9,25]	–	[9,25]	[13]

$$\Delta E = \hat{m}_F(E_d^e + E_d^p + E_d^n) + \tilde{m}_F(E_q^e + E_q^p + E_q^n) \tag{5}$$

where E_d and E_q are energies given below, the superscripts e, p, n stand for electron, proton and neutron and $\hat{m}_F$ and $\tilde{m}_F$ are defined as

$$\hat{m}_F := \frac{m_F}{F}, \qquad \tilde{m}_F := \frac{3m_F^2 - F(F+1)}{3F^2 - F(F+1)} \, . \tag{6}$$

In general ΔE of (5) will be time varying as the energies E_d^w, E_q^w (w stands for e, p, n) depend on the orientation of the angular momentum of w with respect to the fixed stars (best approximation to the frame in which symmetry breaking took place in the early universe). Of particular interest will be (see Sect. 4) Zeeman sublevels with $m_F \neq 0$ in which case the orientation of the quantization axis (quantization magnetic field) is relevant, so one can expect variations of ΔE at sidereal and semi-sidereal frequencies due to the precession of the quantization axis with the rotation of the Earth.

The energies in (5) are [30]

$$\begin{aligned} E_d^w &= \beta_w \tilde{b}_3^w + \delta_w \tilde{d}_3^w + \kappa_w \tilde{g}_d^w \\ E_q^w &= \gamma_w \tilde{c}_q^w + \lambda_w \tilde{g}_q^w \, . \end{aligned} \tag{7}$$

In (7) the tilde quantities have the dimensions of energy and represent laboratory frame combinations of the SME parameters that need to be determined by experiment. They are time varying at sidereal and semi-sidereal frequencies as they are obtained by transforming the constant sun-centered-frame parameters to the laboratory frame. The other coefficients in (7) are constant and depend on the nuclear and electronic structure of the atom. Explicit expressions can be found in [30], with their values calculated for certain atoms and states (including the ^{133}Cs atom of interest to our experiment) in [31].

Our experiment (see Sect. 4) is sensitive to $\tilde{c}_q^p$. When transforming to the sun-centered-frame this parameter is a time varying combination of 8 constant SME parameters ($\tilde{c}_Q$, $\tilde{c}_-$, $\tilde{c}_X$, $\tilde{c}_Y$, $\tilde{c}_Z$, $\tilde{c}_{TX}$, $\tilde{c}_{TY}$, $\tilde{c}_{TZ}$), which are generally used [30,31] to state and compare experimental results (see Table 3). In some publications [42, 43] the results are stated in terms of dimensionless sun-frame parameters related to the $\tilde{c}$ parameters by (c.f. [31] Appendix B)

Table 3. Orders of magnitude of present limits (in GeV) on Lorentz violating parameters in the SME matter sector and corresponding references. Expected uncertainties from the experiment reported in Sect. 4 are given in brackets

Parameter	Proton	Neutron	Electron	References
$\tilde{b}_X$, $\tilde{b}_Y$	10^{-27}	10^{-31}	10^{-29}	[44–46]
$\tilde{b}_Z$	-	-	10^{-28}	[46]
$\tilde{b}_T$, $\tilde{g}_T$, $\tilde{H}_{JT}$, $\tilde{d}_{\pm}$, $\tilde{d}_Q$, $\tilde{d}_{XY}$, $\tilde{d}_{YZ}$	-	10^{-27}	-	[41]
$\tilde{d}_{XZ}$	-	-	-	
$\tilde{d}_X$, $\tilde{d}_Y$	10^{-25}	10^{-29}	10^{-22}	[30, 30, 41]
$\tilde{d}_Z$	-	-	-	
$\tilde{g}_{DX}$, $\tilde{g}_{DY}$	10^{-25}	10^{-29}	10^{-22}	[30, 30, 41]
$\tilde{g}_{DZ}$	-	-	-	
$\tilde{g}_{JK}$	-	-	-	
$\tilde{g}_c$	-	10^{-27}	-	[41]
$\tilde{g}_-$, $\tilde{g}_Q$, $\tilde{g}_{TJ}$	-	-	-	
$\tilde{c}_-$	(10^{-25})	10^{-27}	10^{-19}	[9, 25, 30, 42]
$\tilde{c}_Q$	(10^{-25}) 10^{-11}	-	10^{-9}	[14, 43]
$\tilde{c}_X$, $\tilde{c}_Y$	(10^{-25})	10^{-25}	10^{-19}	[9, 25, 30, 42]
$\tilde{c}_Z$	(10^{-25})	10^{-27}	10^{-19}	[9, 25, 30, 42]
$\tilde{c}_{TJ}$	(10^{-19}) 10^{-8}	-	10^{-6}	[14, 43]

$$\begin{aligned} \tilde{c}_Q &= mc^2(c_{XX} + c_{YY} - 2c_{ZZ}) \\ \tilde{c}_- &= mc^2(c_{XX} - c_{YY}) \\ \tilde{c}_J &= mc^2|\epsilon_{JKL}|c_{KL} \\ \tilde{c}_{TJ} &= mc^2(c_{TJ} + c_{JT}) \end{aligned} \tag{8}$$

where m is the mass of the particle (m_n, m_p, or m_e), indices J, K, L run over sun-frame spatial coordinates X, Y, Z and the totally antisymmetric tensor ϵ_{JKL} is defined with $\epsilon_{XYZ} = +1$.

Existing bounds on the 40 parameters for each particle (n,p,e) come from clock comparison and magnetometer experiments using different atomic species ([31] and references therein, [41]), from resonator experiments (including our experiments described in Sect. 3 as analyzed recently by Müller) [9, 25, 42], and from analysis of Ives-Stilwell (Doppler-shift) experiments [14, 43]. They are summarized in Table 3 below. The expected results of our present experiment (see Sect. 4) are given in Table 3 in brackets. They correspond to first measurements

of some parameters and an improvement by 11 and 14 orders of magnitude on others.

3 Michelson-Morley and Kennedy-Thorndike Tests

In this section we review the results [7–9] of our experiment that compares the frequencies of a cryogenic sapphire oscillator (CSO) and a hydrogen maser atomic clock. Both devices operate at microwave frequencies and are run and compared continuously for timekeeping purposes at the Paris observatory. We use that data to carry out Michelson-Morley and Kennedy-Thorndike experiments, searching for a dependence of the difference frequency on the orientation and/or the velocity of the CSO with respect to a prefered frame candidate.

The heart of the experiment is a monolithic sapphire crystal of cylindrical shape, about 5 cm diameter and 3 cm height. The resonance frequency is determined by exciting a so called Whispering Gallery mode, corresponding to a standing wave set up around the perimeter of the cylinder (see Fig. 1 and [8] for a detailed description). In our case the excited mode is a TE mode at 11.932 GHz, with dominant radial electric and vertical magnetic fields corresponding to propagation (Poynting) vectors in both directions around the circumference. The CSO is an active system oscillating at the resonant frequency (i.e. a classical loop oscillator which amplifies and re-injects the "natural" resonator signal). Additionally the signal is locked to the resonance using the Pound-Drever technique (modulation at ≈80 kHz). The incident power is stabilized in the cryogenic environment and the spurious AM modulation is minimized using a servo loop. To minimize temperature sensitivity the resonator is heated (inside the 4 K environment) and stabilized to the temperature turning point (≈6 K) of the resonator frequency which arises due to paramagnetic impurities in the sapphire. Under these conditions the loaded quality factor of the resonator is slightly below 10^9.

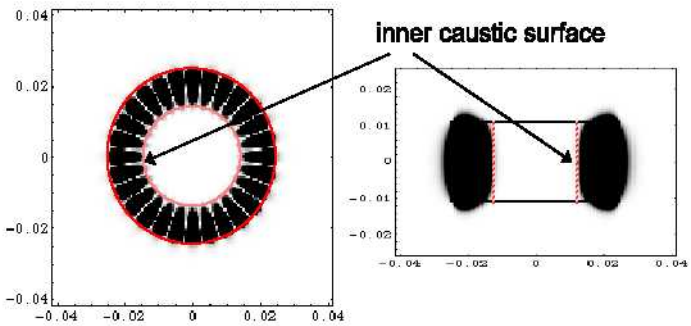


Fig. 1. Magnitude of the H_z field calculated for the $WG_{14,0,0}$ mode in a sapphire disk resonator. The r-ϕ and r-z planes are represented. The inner caustic is shown, and the mode can be approximated as two guided waves propagating in opposite directions around the azimuth

The resonator is kept permanently at cryogenic temperatures, with helium refills taking place about every 20–25 days.

The CSO is compared to a commercial (Datum Inc.) active hydrogen maser whose frequency is also regularly compared to caesium and rubidium atomic fountain clocks in the laboratory [32]. The CSO resonant frequency at 11.932 GHz is compared to the 100 MHz output of the hydrogen maser. The maser signal is multiplied up to 12 GHz of which the CSO signal is subtracted. The remaining ≈67 MHz signal is mixed to a synthesizer signal at the same frequency and the low frequency beat at ≈64 Hz is counted, giving access to the frequency difference between the maser and the CSO. The instability of the comparison chain has been measured at $\leq 2 \times 10^{-14}\tau^{-1}$, with long term instabilities dominated by temperature variations, but not exceeding 10^{-16}.

3.1 Results in the RMS Framework

In the RMS framework our experiment sets the most stringent limit for Kennedy-Thorndike experiments (improving by a factor 70 over previous results) and is among the most precise Michelson-Morley tests (see Table 1). Those results were reported in [7,8] and are summarized here.

In the RMS framework the frequency of a resonator in the lab frame S is proportional to ${t_c}^{-1}$ where t_c is the return travel time of a light signal in the resonator. Setting $c^2dT^2 = dX^2 + dY^2 + dZ^2$ in the preferred frame Σ, and transforming according to (1) we find the coordinate travel time of a light signal in S:

$$dt = \frac{dl}{c}\left(1 - (\beta_{\rm MS} - \alpha_{\rm MS} - 1)\,\frac{v^2}{c^2} - \left(\frac{1}{2} - \beta_{\rm MS} + \delta_{\rm MS}\right)\sin^2\theta\frac{v^2}{c^2}\right) + \mathcal{O}(4) \quad (9)$$

where $dl = \sqrt{dx^2 + dy^2 + dz^2}$ and θ is the angle between the direction of light propagation and the velocity $\mathbf{v}$ of S in Σ.

Calculating t_c from (9) the relative frequency difference between the sapphire oscillator and the hydrogen maser (which, by definition, realizes coordinate time in S [33]) is

$$\frac{\Delta\nu(t)}{\nu_0} = P_{KT}\frac{v(t)^2}{c^2} + P_{MM}\frac{v(t)^2}{c^2}\frac{1}{2\pi}\int_0^{2\pi}\sin^2\theta(t,\varphi)d\varphi + \mathcal{O}(3) \quad (10)$$

where ν_0 is the unperturbed frequency, $v(t)$ is the (time dependent) speed of the lab in Σ, and φ is the azimuthal angle of the light signal in the plane of the cylinder. The periodic time dependence of v and θ due to the rotation and orbital motion of the Earth with respect to the CMB frame allow us to set limits on the two parameters in (10) by fitting the periodic terms of appropriate frequency and phase (see [35] for calculations of similar effects for several types of oscillator modes). Given the limited durations of our data sets (≤16 days) the dominant periodic terms arise from the Earth's rotation, so retaining only those we have $\mathbf{v}(t) = \mathbf{u} + \omega \times \mathbf{R}$ with $\mathbf{u}$ the velocity of the solar system with respect to the

CMB, ω the angular velocity of the Earth, and $\mathbf{R}$ the geocentric position of the lab. We then find after some calculation.

$$\frac{\Delta\nu}{\nu_0} = P_{KT}(H\sin\lambda) \qquad (11)$$
$$+P_{MM}(A\cos\lambda + B\cos(2\lambda) + C\sin\lambda + D\sin\lambda\cos\lambda + E\sin\lambda\cos(2\lambda))$$

where $\lambda = \omega t + \phi$, and A to E and ϕ are constants depending on the latitude and longitude of the lab ($\approx$48.7°N and 2.33°E for Paris). Numerically $H \approx -2.6 \times 10^{-9}$, $A \approx -8.8 \times 10^{-8}$, $B \approx 1.8 \times 10^{-7}$, C to E of order 10^{-9}. We note that in (11) the dominant time variations of the two combinations of parameters are in quadrature and at twice the frequency which indicates that they should decorelate well in the data analysis allowing a simultaneous determination of the two (as confirmed by the correlation coefficients given below). Fitting this simplified model to our data we obtain results that differ by less than 10% from the results presented below that were obtained using the complete model ((10) including the orbital motion of the Earth).

For the RMS analysis we use 13 data sets in total spanning Sept. 2002 to Aug. 2003, of differing lengths (5 to 16 days, 140 days in total). The sampling time for all data sets was 100 s except two data sets with $\tau_0 = 12$ s. To make the data more manageable we first average all points to $\tau_0 = 2500$ s. For the data analysis we simultaneously fit (using weighted least squares, WLS, c.f. [8]) an offset and a rate (natural frequency drift, typically $\approx 1.7 \times 10^{-18}$ s^{-1}) per data set and the two parameters of the model (10). In the model (10) we take into account the rotation of the Earth and the Earth's orbital motion, the latter contributing little as any constant or linear terms over the durations of the individual data sets are absorbed by the fitted offsets and rates.

Figure 2 shows the resulting values of the two parameters (P_{KT} and P_{MM}) for each individual data set. A global WLS fit of the two parameters and the 13 offsets and drifts yields $P_{MM} = (1.2 \pm 1.9) \times 10^{-9}$ and $P_{KT} = (1.6 \pm 2.3) \times 10^{-7}$ (1σ uncertainties), with the correlation coefficient between the two parameters less than 0.01 and all other correlation coefficients $<$0.06. The distribution of the 13 individual values around the ones obtained from the global fit is well compatible with a normal distribution ($\chi^2 = 10.7$ and $\chi^2 = 14.6$ for P_{MM} and P_{KT} respectively).

Systematic effects at diurnal or semi-diurnal frequencies with the appropriate phase could mask a putative sidereal signal. The statistical uncertainties of P_{MM} and P_{KT} obtained from the WLS fit above correspond to sidereal and semi-sidereal terms (from (11)) of $\approx 7 \times 10^{-16}$ and $\approx 4 \times 10^{-16}$ respectively so any systematic effects exceeding these limits need to be taken into account in the final uncertainty. We expect the main contributions to such effects to arise from temperature, pressure and magnetic field variations that would affect the hydrogen maser, the CSO and the associated electronics, and from tilt variations of the CSO which are known to affect its frequency (see Sect. 3.2 for a detailed discussion). Our final uncertainties (the error bars in Fig. 2) are the quadratic sums of the statistical uncertainties from the WLS adjustment for each data set

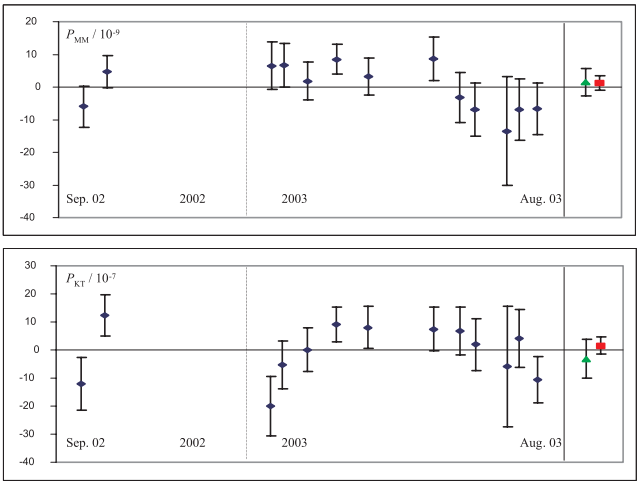


Fig. 2. Values (published in [8]) of the two parameters (P_{KT} and P_{MM}) from a fit to each individual data set (*blue diamonds*) and a global fit to all the data (*red squares*). For comparison the previous results published in [7] are also shown (*green triangles*). The error bars indicate the combined uncertainties from statistics and systematic effects

and the systematic uncertainties calculated for each data set from (11). For the global adjustment we average the systematic uncertainties from the individual data sets obtaining $\pm 1.2 \times 10^{-9}$ on P_{MM} and $\pm 1.9 \times 10^{-7}$ on P_{KT}.

In the RMS framework, our experiment simultaneously constrains two combinations of the three parameters of the Mansouri and Sexl test theory (previously measured individually by Michelson-Morley and Kennedy-Thorndike experiments). We obtain $\delta_{\rm MS} - \beta_{\rm MS} + 1/2 = 1.2(1.9)(1.2) \times 10^{-9}$ which is of the same order as the best previous results [23,25], and $\beta_{\rm MS} - \alpha_{\rm MS} - 1 = 1.6(2.3)(1.9) \times 10^{-7}$ which improves the best previous limit [27] by a factor of 70 (the first bracket indicates the 1σ uncertainty from statistics the second from systematic effects). We note that our value on $\delta_{\rm MS} - \beta_{\rm MS} + 1/2$ is compatible with the slightly significant recent result of [25] who obtained $\delta_{\rm MS} - \beta_{\rm MS} + 1/2 = (2.2 \pm 1.5) \times 10^{-9}$.

As a result of our experiment the Lorentz transformations are confirmed in the RMS framework (c.f. Table 1) with an overall uncertainty of $\leq 3 \times 10^{-7}$ limited by our determination of $\beta_{\rm MS} - \alpha_{\rm MS} - 1$ and the recent limit [14] of 2.2×10^{-7} on the determination of $\alpha_{\rm MS}$. The latter is likely to improve in the coming years by experiments such as ACES (Atomic Clock Ensemble in Space [36]) that will compare ground clocks to clocks on the international space station aiming at a 10^{-8} measurement of $\alpha_{\rm MS}$.

3.2 Results in the SME

In the SME our experiment sets the presently most stringent limits on a number of photon sector parameters, improving previous results [25] by up to an order of magnitude. These results were first published in [9] and are reproduced here.

The SME perturbed frequency of a resonator can be calculated from equation (3) in the form (c.f. [19])

$$\frac{\Delta\nu}{\nu_0} = -\frac{1}{\langle U\rangle}\int_V d^3x\left(\epsilon_0 \mathbf{E_0}^*\cdot\kappa_{DE}\cdot\mathbf{E_0} - \mu_0^{-1}\mathbf{B_0}^*\cdot\kappa_{HB}\cdot\mathbf{B_0} + 2\mathrm{Re}\left(\sqrt{\frac{\epsilon_0}{\mu_0}}\mathbf{E_0}^*\cdot\kappa_{DB}\cdot\mathbf{B_0}\right)\right) \tag{12}$$

where $\mathbf{B_0}, \mathbf{H_0}, \mathbf{E_0}, \mathbf{D_0}$ are the unperturbed (standard Maxwell) fields and $\langle U\rangle = \int_V d^3x(\mathbf{E_0}\cdot\mathbf{D_0}^* + \mathbf{B_0}\cdot\mathbf{H_0}^*)$. Note that, as shown in [30], the frequency of the H-maser is not affected to first order (because it operates on $m_F = 0$ states) and [37] shows that the perturbation of the frequency due to the modification of the sapphire crystal structure (and hence the cavity size) is negligible with respect to the direct perturbation of the e-m fields.

The resonator is placed in the lab with its symmetry axis along the vertical. Applying (12) in the lab frame (z-axis vertical upwards, x-axis pointing south), with the fields calculated using a finite element technique as described in [8], we obtain an expression for the frequency variation of the resonator

$$\begin{aligned}\frac{\Delta\nu}{\nu_0} = {} & (\mathcal{M}_{DE})^{xx}_{lab}\left((\kappa_{DE})^{xx}_{lab} + (\kappa_{DE})^{yy}_{lab}\right) + (\mathcal{M}_{DE})^{zz}_{lab}(\kappa_{DE})^{zz}_{lab} \\ & + (\mathcal{M}_{HB})^{xx}_{lab}\left((\kappa_{HB})^{xx}_{lab} + (\kappa_{HB})^{yy}_{lab}\right) + (\mathcal{M}_{HB})^{zz}_{lab}(\kappa_{HB})^{zz}_{lab}\end{aligned} \tag{13}$$

with the $\mathcal{M}_{lab}$ components given in Table 4. To obtain the values in Table 4 we take into account the fields inside the resonator (c.f. [8]) and outside ($\leq$2% of the energy).

Table 4. $\mathcal{M}_{lab}$ components calculated using (12) and a finite element technique for the determination of the fields inside the resonator (see [8] for details)

$(\mathcal{M}_{DE})^{xx}_{lab}$	$(\mathcal{M}_{DE})^{zz}_{lab}$	$(\mathcal{M}_{HB})^{xx}_{lab}$	$(\mathcal{M}_{HB})^{zz}_{lab}$
−0.03093	−0.0004030	0.008408	0.4832

The last step is to transform the κ tensors in (13) to the conventional sun-centered frame using the explicit transformations provided in [19], and to express the result in terms of the $\tilde{\kappa}$ tensors of (4). We obtain

$$\frac{\nu-\nu_0}{\nu_0} = \sum_i C_i\cos(\omega_i T_\oplus + \varphi_i) + S_i\sin(\omega_i T_\oplus + \varphi_i) \tag{14}$$

Table 5. Coefficients C_i and S_i in (1) for the six frequencies ω_i of interest and their relation to the components of the SME parameters $\tilde{\kappa}_{e-}$ and $\tilde{\kappa}_{o+}$, with $\omega_\oplus$ and $\Omega_\oplus$ the angular frequencies of the Earth's sidereal rotation and orbital motion. The measured values (in 10^{-16}) are shown together with the statistical (first bracket) and systematic (second bracket) uncertainties.

ω_i	C_i	S_i
$\omega_\oplus - \Omega_\oplus$	$(-8.6\times10^{-6})\tilde{\kappa}_{o+}^{YZ}$	$(8.6\times10^{-6})\tilde{\kappa}_{o+}^{XZ} - (4.2\times10^{-5})\tilde{\kappa}_{o+}^{XY}$
$\omega_\oplus$	$-0.44\tilde{\kappa}_{e-}^{XZ} + (1.1\times10^{-6})\tilde{\kappa}_{o+}^{XZ}$	$-0.44\tilde{\kappa}_{e-}^{YZ} + (1.1\times10^{-6})\tilde{\kappa}_{o+}^{YZ}$
$\omega_\oplus + \Omega_\oplus$	$(-8.6\times10^{-6})\tilde{\kappa}_{o+}^{YZ}$	$(8.6\times10^{-6})\tilde{\kappa}_{o+}^{XZ} + (1.8\times10^{-6})\tilde{\kappa}_{o+}^{XY}$
$2\omega_\oplus - \Omega_\oplus$	$(-1.8\times10^{-5})\tilde{\kappa}_{o+}^{XZ}$	$(-1.8\times10^{-5})\tilde{\kappa}_{o+}^{YZ}$
$2\omega_\oplus$	$-0.10(\tilde{\kappa}_{e-}^{XX} - \tilde{\kappa}_{e-}^{YY})$	$-0.19\tilde{\kappa}_{e-}^{XY}$
$2\omega_\oplus + \Omega_\oplus$	$(7.8\times10^{-7})\tilde{\kappa}_{o+}^{XZ}$	$(7.8\times10^{-7})\tilde{\kappa}_{o+}^{YZ}$
$\omega_\oplus - \Omega_\oplus$	−6.9(4.2)(4.5)	6.7(4.2)(4.5)
$\omega_\oplus$	14(4.2)(4.2)	2.4(4.2)(4.2)
$\omega_\oplus + \Omega_\oplus$	−6.0(4.2)(4.2)	2.7(4.2)(4.2)
$2\omega_\oplus - \Omega_\oplus$	3.7(2.4)(3.7)	−2.9(2.4)(3.7)
$2\omega_\oplus$	3.1(2.4)(3.7)	11(2.4)(3.7)
$2\omega_\oplus + \Omega_\oplus$	0.0(2.4)(3.7)	−1.2(2.4)(3.7)

where ν_0 is the unperturbed frequency difference, the sum is over the six frequencies ω_i of Table 5, the coefficients C_i and S_i are functions of the Lorentz violating tensors $\tilde{\kappa}_{e-}$ and $\tilde{\kappa}_{o+}$ (see Table 5), $T_\oplus = 0$ on December 17, 2001, 18:05:16 UTC, $\varphi_{\omega_\oplus} = \varphi_{2\omega_\oplus} = 0$ and $\varphi_{(\omega_\oplus \pm \Omega_\oplus)} = \varphi_{(2\omega_\oplus \pm \Omega_\oplus)} = \pm 4.682$ rad. To obtain the relations of Table 5 between C_i, S_i and the SME parameters we have assumed zero values for the 10 independent components of the $\tilde{\kappa}_{e+}$ and $\tilde{\kappa}_{o-}$ tensors, as those have been determined to $\leq 2\times10^{-32}$ by astrophysical tests [19].

To determine all 7 SME parameters appearing in Table 5 one requires over a year of data in order to be able to decorrelate the annual sidebands from the sidereal and twice sidereal frequencies. To do so we have extended the data to 20 data sets in total, spanning Sept. 2002 to Jan. 2004, of differing lengths (5 to 20 days, 222 days in total). The sampling time for all data sets was 100 s.

For the statistical analysis we first average the data to 2500 s sampling time and then simultaneously fit the 20 rates and offsets and the 12 parameters C_i and S_i of (14) to the complete data using two statistical methods, weighted least squares (WLS), which allows one to account for non-white noise processes (cf. [7]), and individual periods (IP) as used in [25]. The two methods give similar results for the parameters (within the uncertainties) but differ in the estimated uncertainties (the IP uncertainties are a factor ≈1.2 larger). Because IP discards a significant amount of data (about 10% in our case) we consider WLS the more realistic method and retain those results as the statistical uncertainties

shown in Table 5. We note that we now have sufficient data to decorrelate all 12 parameters (C_i, S_i) i.e. the WLS correlation coefficients between any two parameters or between any parameter and the fitted offsets and rates are all less than 0.20.

To investigate the distributions of our results we fit the coefficients C_i and S_i to each one of the 20 data sets individually with the results at the sidereal and semi-sidereal frequencies $\omega_\oplus$ and $2\omega_\oplus$ shown in Fig. 3. If a genuine effect at those frequencies was present we would expect correlated phases of the individual points in Fig. 3, but this does not seem to be supported by the data. A distribution of the phases may result from an effect at a neighboring frequency, in particular the diurnal and semi-diurnal frequencies $\omega_\oplus - \Omega_\oplus$ and $2(\omega_\oplus - \Omega_\oplus)$ at which we would expect systematic effects to play an important role. Figure 4 shows the amplitudes $A_\omega = \sqrt{C_\omega^2 + S_\omega^2}$ resulting from least squares fits for a range of frequencies, ω, around the frequencies of interest. We note that the fitted amplitudes at $\omega_\oplus - \Omega_\oplus$ and $2(\omega_\oplus - \Omega_\oplus)$ are substantially smaller than those at $\omega_\oplus$ and $2\omega_\oplus$ and therefore unlikely to contribute to the distribution of the points in Fig. 3.

Systematic effects at the frequencies ω_i could mask a putative Lorentz violating signal in our experiment and need to be investigated in order to be able to confirm such a signal or to exclude it within realistic limits. We have extensively studied all systematic effects arising from environmental factors that might affect our experiment. The resulting estimated contributions at the two central frequencies $\omega_\oplus$, $2\omega_\oplus$ and at the diurnal frequency $\omega_\oplus - \Omega_\oplus$ are summarized in Table 6. The contributions at $\omega_\oplus + \Omega_\oplus$ and $2\omega_\oplus \pm \Omega_\oplus$ are not shown as they are identical to those at $\omega_\oplus$ and $2\omega_\oplus$ respectively.

We have compared the Hydrogen-maser (HM) used as our frequency reference to our highly stable and accurate Cs fountain clocks (FO2 and FOM). For example, the amplitudes at $\omega_\oplus$ and $2\omega_\oplus$ of the HM-FOM relative fre-

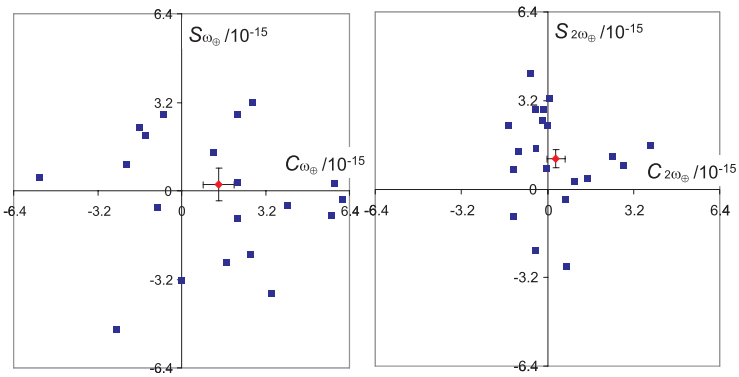


Fig. 3. Fitted sine and cosine amplitudes at $\omega_\oplus$ and $2\omega_\oplus$ for each data set (*blue squares*) and the complete data (*red diamonds*, with statistical errors). For clarity the error bars of the individual data sets have been omitted

Table 6. Contributions from systematic effects to the amplitudes A_i (parts in 10^{16}) at three frequencies ω_i

Effect	$\omega_\oplus - \Omega_\oplus$	$\omega_\oplus$	$2\omega_\oplus$
H-maser	< 5	< 5	< 5
Tilt	3	3	1
Gravity	0.3	0.3	0.3
B-field	< 0.1	< 0.1	< 0.1
Temperature	< 1	< 1	< 1
Atm. Pressure	2.3	0.3	0.4
Total	**6.4**	**5.9**	**5.2**

quency difference over June-July 2003 were $A_{\omega_\oplus} = (4.8 \pm 4.7) \times 10^{-16}$ and $A_{2\omega_\oplus} = (4.3 \pm 4.7) \times 10^{-16}$. This indicates that any environmental effects on the HM at those frequencies should be below 5 parts in 10^{16} in amplitude. This is in good agreement with studies on similar HMs carried out in [38] that limited environmental effects to < 3 to 4 parts in 10^{16}.

To estimate the tilt sensitivity we have intentionally tilted the oscillator by ≈5 mrad off its average position which led to relative frequency variations of $\approx 3 \times 10^{-13}$ from which we deduce a tilt sensitivity of $\approx 6 \times 10^{-17} \mu\text{rad}^{-1}$. This is in good agreement with similar measurements in [39] that obtained sensitivities of $\approx 4 \times 10^{-17} \mu\text{rad}^{-1}$. Measured tilt variations in the lab at diurnal and semi-diurnal periods show amplitudes of 4.6 μrad and 1.6 μrad respectively which leads to frequency variations that do not exceed 3×10^{-16} and 1×10^{-16} respectively.

From the measurements of tilt sensitivity one can deduce the sensitivity to gravity variations (cf. [39]), which in our case lead to a sensitivity of $\approx 3 \times 10^{-10} g^{-1}$. Tidal gravity variations can reach $\approx 10^{-7} g$ from which we obtain a

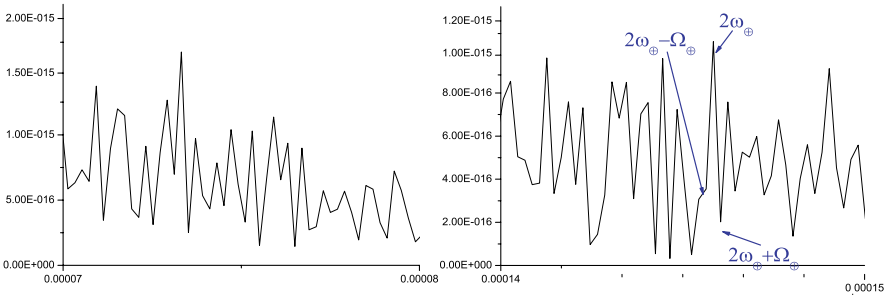


Fig. 4. Fitted Amplitudes A_ω for a range of frequencies around the six frequencies ω_i of interest (indicated by *arrows*)

maximum effect of 3×10^{-17}, one order of magnitude below the effect from tilt variations.

Variations of the ambient magnetic field in our lab. are dominated by the passage of the Paris Metro, showing a strong periodicity ("quiet" periods from 1 am to 5 am). The corresponding diurnal and semi-diurnal amplitudes are 1.7×10^{-4} G and 3.4×10^{-4} G respectively for the vertical field component and about 10 times less for the horizontal one. To determine the magnetic sensitivity of the CSO we have applied a sinusoidal vertical field of 0.1 G amplitude with a 200 s period. Comparing the CSO frequency to the FO2 Cs-fountain we see a clear sinusoidal signal (S/N > 2) at the same period with an amplitude of 7.2×10^{-16}, which leads to a sensitivity of $\approx 7 \times 10^{-15}$ G^{-1}. Assuming a linear dependence (there is no magnetic shielding that could lead to non-linear effects) we obtain effects of only a few parts in 10^{-18}.

Late 2002 we implemented an active temperature stabilization inside an isolated volume ($\approx 15\text{m}^3$) that includes the CSO and all the associated electronics. The temperature is measured continously in two fixed locations (behind the electronics rack and on top of the dewar). For the best data sets the measured temperature variations do not exceed 0.02/0.01 K in amplitude for the diurnal and semi-diurnal components. A least squares fit to all our temperature data (taken simultaneously with our frequency measurements) yields amplitudes of $A_{\omega_\oplus} = 0.020$ K and $A_{2\omega_\oplus} = 0.018$ K with similar values at the other frequencies ω_i of interest, including the diurnal one ($A_{\omega_\oplus - \Omega_\oplus} = 0.022$ K). Inducing a strong sinusoidal temperature variation (≈ 0.5 K amplitude at 12 h period) leads to no clearly visible effect on the CSO frequency. Taking the noise level around the 12 h period as the maximum effect we obtain a sensitivity of $< 4 \times 10^{-15}$ per K. Using this estimate we obtain effects of $< 1 \times 10^{-16}$ at all frequencies ω_i.

Finally we have investigated the sensitivity of the CSO to atmospheric pressure variations. To do so we control the pressure inside the dewar using a variable valve mounted on the He-gas exhaust. During normal operation the valve is open and the CSO operates at ambient atmospheric pressure. For the sensitivity determination we have induced a sinusoidal pressure variation (≈ 14 mbar amplitude at 12 h period), which resulted in a clearly visible effect on the CSO frequency corresponding to a sensitivity of $\approx 6.5 \times 10^{-16}$ mbar^{-1}. We have checked that the sensitivity is not significantly affected when changing the amplitude of the induced pressure variation by a factor 3. A least squares fit to atmospheric pressure data (taken simultaneously with our frequency measurements) yields amplitudes of $A_{\omega_\oplus} = 0.045$ mbar and $A_{2\omega_\oplus} = 0.054$ mbar with similar values at the other frequencies ω_i of interest, except the diurnal one for which $A_{\omega_\oplus - \Omega_\oplus} = 0.36$ mbar. The resulting effects on the CSO frequency are given in Table 6.

Our final results for the 7 components of $\tilde{\kappa}_{e-}$ and $\tilde{\kappa}_{o+}$ are obtained from a least squares fit to the 12 measured coefficients of Table 5. They are summarized and compared to the results of [25] in Table 7.

We note that our results for $\tilde{\kappa}_{e-}^{XY}$ and $\tilde{\kappa}_{e-}^{XZ}$ are significant at about 2σ, while those of [25] are significant at about the same level for $(\tilde{\kappa}_{e-}^{XX} - \tilde{\kappa}_{e-}^{YY})$. The two experiments give compatible results for $\tilde{\kappa}_{e-}^{XZ}$ (within the 1σ uncertainties) but

Table 7. Results for the components of the SME Lorentz violation parameters $\tilde{\kappa}_{e-}$ (in 10^{-15}) and $\tilde{\kappa}_{o+}$ (in 10^{-11})

	$\tilde{\kappa}_{e-}^{XY}$	$\tilde{\kappa}_{e-}^{XZ}$	$\tilde{\kappa}_{e-}^{YZ}$	$(\tilde{\kappa}_{e-}^{XX} - \tilde{\kappa}_{e-}^{YY})$
from [25]	1.7(2.6)	−6.3(12.4)	3.6(9.0)	8.9(4.9)
this work	−5.7(2.3)	−3.2(1.3)	−0.5(1.3)	−3.2(4.6)
	$\tilde{\kappa}_{o+}^{XY}$	$\tilde{\kappa}_{o+}^{XZ}$	$\tilde{\kappa}_{o+}^{YZ}$	
from [25]	14(14)	−1.2(2.6)	0.1(2.7)	
this work	−1.8(1.5)	−1.4(2.3)	2.7(2.2)	

not for the other two parameters, so the measured values of those are unlikely to come from a common source. Another indication for a non-genuine effect comes from Figs. 3 and 4, as we would expect any genuine effect to show an approximately coherent phase for the individual data sets in Fig. 3 and to display more prominent peaks in Fig. 4.

In conclusion, we have not seen any Lorentz violating effects in the general framework of the SME, and set limits on 7 parameters of the SME photon sector (cf. Table 7) which are up to an order of magnitude more stringent than those obtained from previous experiments [25]. Two of the parameters are significant (at $\approx 2\sigma$). We believe that this is most likely a statistical coincidence or a neglected systematic effect. To verify this, our experiment is continuing and new, more precise experiments are under way [35].

4 Atomic Clock Test of Lorentz Invariance in the SME Matter Sector

For this experiment we use one of the laser cooled fountain clocks operated at the Paris observatory, the ^{133}Cs and ^{87}Rb double fountain FO2 [47]. We run it in Cs mode on the $|F = 4\rangle \leftrightarrow |F = 3\rangle$ hyperfine transition of the $6S_{1/2}$ ground state. Both hyperfine states are degenerate, with Zeeman substates $m_F = [-4, 4]$ and $m_F = [-3, 3]$ respectively. The clock transition used in routine operation is $|F = 4, m_F = 0\rangle \leftrightarrow |F = 3, m_F = 0\rangle$ at 9.2 GHz, which is magnetic field independent to first order. The first order magnetic field dependent Zeeman transitions ($|F = 4, m_F = i\rangle \leftrightarrow |F = 3, m_F = i\rangle$ with $i = \pm 1, \pm 2, \pm 3$) are used regularly for measurement and characterization of the magnetic field, necessary to correct the second order Zeeman effect of the clock transition. In routine operation the clock transition frequency stability of FO2 is $1.6 \times 10^{-14}\tau^{-1/2}$, and its accuracy 7×10^{-16} [47,48], the best performance of any clock at present.

In the presence of Lorentz violation the SME frequency shift of a Cs $|F = 4, m_F\rangle \leftrightarrow |F = 3, m_F\rangle$ transition, arising from the energy level shifts described in Sect. 2.2, has been calculated explicitly in [31]. It can be written in the form

Table 8. Coefficients entering (15) for a $^{133}Cs\ |F=4,m_F\rangle \leftrightarrow |F=3,m_F\rangle$ transition. $K_p = \langle p^2\rangle/m_p^2$ for the Schmidt proton and $K_e = \langle p^2\rangle/m_e^2$ for the valence electron, with $K_p \approx 10^{-2}$ and $K_e \approx 10^{-5}$ [31]

β_p	δ_p	κ_p	γ_p	λ_p	β_e	δ_e	κ_e	s_1^p	s_2^p	s_1^e
$\frac{7}{9}$	$-\frac{7}{33}K_p$	$\frac{28}{99}K_p$	$-\frac{1}{9}K_p$	0	-1	$\frac{1}{3}K_e$	$-\frac{1}{3}K_e$	$-\frac{1}{14}m_F$	$-\frac{1}{14}m_F^2$	$\frac{1}{2}m_F$

$$\hbar(\delta\omega_{SME}) = s_1^p\left(\beta_p\tilde{b}_3^p - \delta_p\tilde{d}_3^p + \kappa_p\tilde{g}_d^p\right) + s_2^p\left(\gamma_p\tilde{c}_q^p - \lambda_p\tilde{g}_q^p\right) + s_1^e\left(\beta_e\tilde{b}_3^e - \delta_e\tilde{d}_3^e + \kappa_e\tilde{g}_d^e\right) \tag{15}$$

where the tilde quantities are the SME matter sector parameters described in Sect. 2.2. The quantities $\beta_w, \delta_w, \kappa_w, \gamma_w, \lambda_w$ depend on the nuclear and electronic structure, and are given in table II of [31]. The s coefficients result from the application of the Wigner-Eckhart theorem and are also given in [31]. All coefficients entering (15) are summarized in Table 8.

From (15) and Table 8 we notice that all $m_F \neq 0$ Zeeman transitions are sensitive to a violation of Lorentz symmetry, but not the $m_F = 0$ clock transition. So in principle a direct measurement of one of the Zeeman transitions with respect to the clock transition (used as the reference) can yield a test of Lorentz invariance. The sensitive axis of the experiment is defined by the direction of the quantization magnetic field used to separate the Zeeman substates (vertical in the case of FO2), hence the rotation of the earth provides a modulation of the Lorentz violating signal at sidereal and semi-sidereal frequencies, which could be searched for in the data.

However, in such a direct measurement the first order Zeeman shift of the $m_F \neq 0$ transition would be the dominant error source and largely degrade the sensitivity of the experiment. The complete frequency shift of a Cs hyperfine Zeeman transition is [49]

$$\delta\omega = \delta\omega_{SME} + m_F K_Z^{(1)} B + \left(1 - \frac{m_F^2}{16}\right) K_Z^{(2)} B^2 + \Delta \tag{16}$$

where $\delta\omega_{SME}$ is the SME frequency shift given by (15), B is the magnetic field seen by the atom, $K_Z^{(1)} = 44.035$ rad s^{-1} nT^{-1} is the first order Zeeman coefficient, $K_Z^{(2)} = 2685.75$ rad s^{-1} T^{-2} is the second order coefficient, and Δ is the shift due to other systematic effects. In (16) the diurnal and semi-diurnal variations of B would mimic a putative Lorentz violating signal appearing in the sidereal and semi-sidereal variations of $\delta\omega_{SME}$ and render such a measurement very uncertain.

A somewhat cleverer strategy is to take advantage of the linear dependence on m_F of the first order Zeeman shift but quadratic dependence on m_F of one of the SME terms (the s_2^p term in (15)). That implies that when measuring "simultaneously" the $m_F = 3$, $m_F = -3$, and $m_F = 0$ transitions and forming the observable $(\omega_{+3} + \omega_{-3} - 2\omega_0)$ one should obtain a quantity that is independent

of the first order Zeeman shift, but still shows a deviation from zero and a sidereal and semi-sidereal modulation in the presence of Lorentz violation. Using (15) and (16) this observable is

$$(\omega_{+3} + \omega_{-3} - 2\omega_0) = \frac{1}{7} K_p \tilde{c}_q^p + K_{Z(obs)}^{(2)} B^2 + \Delta_{(obs)} \tag{17}$$

where $K_{Z(obs)}^{(2)}$ and Δ_{obs} are now the second order Zeeman coefficient and correction from other systematic effects for the complete observable.

The first term of (17) characterizes a possible Lorentz violation in the SME and is time varying when transforming the lab frame parameter $\tilde{c}_q^p$ to the conventional sun-centered frame. The general form of that transformation yields [31]

$$\tilde{c}_q^p = \tilde{B} + \tilde{C}_{\omega_\oplus} \cos(\omega_\oplus t) + \tilde{S}_{\omega_\oplus} \sin(\omega_\oplus t) + \tilde{C}_{2\omega_\oplus} \cos(2\omega_\oplus t) + \tilde{S}_{2\omega_\oplus} \sin(2\omega_\oplus t) \tag{18}$$

where $\omega_\oplus$ is the frequency of rotation of the Earth. The coefficients $\tilde{B}$, $\tilde{C}_{\omega_\oplus}$, $\tilde{S}_{\omega_\oplus}$, $\tilde{C}_{2\omega_\oplus}$, $\tilde{S}_{2\omega_\oplus}$ are functions of the 8 constant sun frame SME parameters $\tilde{c}_X^p$, $\tilde{c}_Y^p$, $\tilde{c}_Z^p$, $\tilde{c}_Q^p$, $\tilde{c}_-^p$, $\tilde{c}_{TX}^p$, $\tilde{c}_{TY}^p$, $\tilde{c}_{TZ}^p$ (see [31] for details) with the three $\tilde{c}_{TJ}^p$ components suppressed by a factor $v_R/c \approx 10^{-6}$ related to the velocity v_R of the lab due to the rotation of the Earth.

The observable we use (17) should be independent of any long term (>few seconds) variations of the first order Zeeman effect and therefore any sidereal or semi-sidereal variation of the observable would be the result of Lorentz violation, if it exceeds the measurement noise and the limits imposed by other systematic effects (see below).

The FO2 setup is sketched in Fig. 5. Cs atoms effusing from an oven are slowed using a counter propagating laser beam and captured in a lin ⊥ lin optical molasses. Atoms are cooled by six laser beams supplied by preadjusted fiber

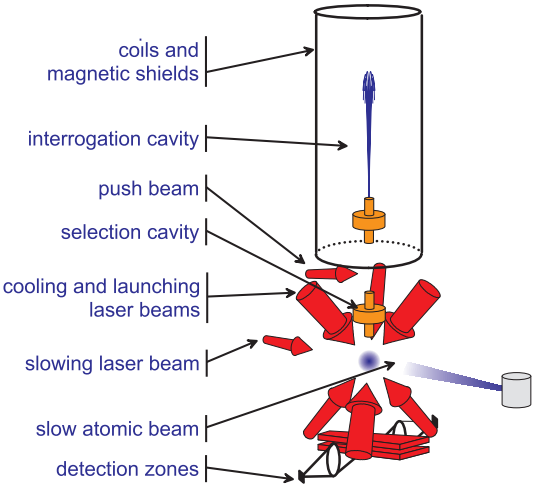


Fig. 5. Schematic view of an atomic fountain

couplers precisely attached to the vacuum tank and aligned along the axes of a 3 dimensional coordinate system, where the (111) direction is vertical. Compared to typical clock operation [47], the number of atoms loaded in the optical molasses has been reduced to 2×10^7 atoms captured in 30 ms.

Atoms are launched upwards at 3.94 m.s^{-1} by using a moving optical molasses and cooled to $\sim$1 μK in the moving frame by adiabatically decreasing the laser intensity and increasing the laser detuning. Atoms are then selected by means of a microwave excitation in the selection cavity performed in a bias magnetic field of $\sim$20 μT, and of a push laser beam. Any of the $|F = 3, m_F\rangle$ states can be prepared with a high degree of purity (few 10^{-3}) by tuning the selection microwave frequency. 52 cm above the capture zone, a cylindrical copper cavity (TE_{011} mode) is used to probe the $|F = 3, m_F\rangle \leftrightarrow |F = 4, m_F\rangle$ hyperfine transition at 9.2 GHz. The Ramsey interrogation method is performed by letting the atomic cloud interact with the microwave field a first time on the way up and a second time on the way down. After the interrogation, the populations $N_{F=4}$ and $N_{F=3}$ of the two hyperfine levels are measured by laser induced fluorescence, leading to a determination of the transition probability $P = N_{F=3}/(N_{F=3} + N_{F=4})$ which is insensitive to atom number fluctuations. One complete fountain cycle from capture to detection lasts 1045 ms in the present experiment. From the transition probability, measured on both sides of the central Ramsey fringe, we compute an error signal to lock the microwave interrogation frequency to the atomic transition using a digital servo loop. The frequency corrections are applied to a computer controlled high resolution DDS synthesizer in the microwave generator. These corrections are used to measure the atomic transition frequency with respect to the local reference signal used to synthesize the microwave frequency.

The homogeneity and the stability of the magnetic field in the interrogation region is a crucial point for the experiment. A magnetic field of 200 nT is produced by a main solenoid (length 815 mm, diameter 220 mm) and a set of 4 compensation coils. These coils are surrounded by a first layer of 3 cylindrical magnetic shields. A second layer is composed of 2 magnetic shields surrounding the entire experiment (optical molasses and detection zone included). Between the two layers, the magnetic field fluctuations are sensed with a flux-gate magnetometer and stabilized by acting on 4 hexagonal coils. The magnetic field in the interrogation region is probed using the $|F = 3, m_F = 1\rangle \leftrightarrow |F = 4, m_F = 1\rangle$ atomic transition with a sensitivity of 7.0084 Hz.nT^{-1}. Measurements of the transition frequency as a function of the launch height show a peak to peak spatial variation of 230 pT over a range of 320 mm above the interrogation cavity. Measurements of the same transition as a function of time at the launch height of 791 mm show a magnetic field instability near 2 pT at τ =1 s as indicated in Fig. 6. The long term behavior exhibits residual variations of the magnetic field induced by temperature fluctuations which could cause variations of the current flowing through solenoid, of the solenoid geometry, of residual thermoelectric currents, of the magnetic shield permeability, etc.

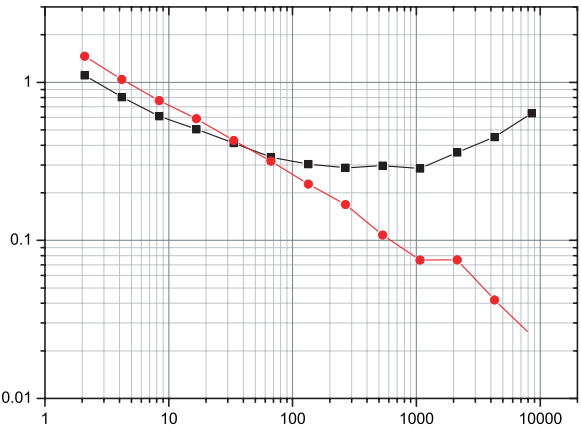


Fig. 6. Magnetic field instability as a function of integration time τ

The experimental sequence is tailored to circumvent the limitation that the long term magnetic field fluctuations could cause. First $|F = 3, m_F = -3\rangle$ atoms are selected and the $|F = 3, m_F = -3\rangle \leftrightarrow |F = 4, m_F = -3\rangle$ transition is probed at half maximum on the red side of the resonance (0.528 Hz below the resonance center). The next fountain cycle, $|F = 3, m_F = +3\rangle$ atoms are selected and the $|F = 3, m_F = +3\rangle \leftrightarrow |F = 4, m_F = +3\rangle$ transition is also probed at half maximum on the red side of the resonance. The third fountain cycle, $|F = 3, m_F = -3\rangle$ atoms are selected and the $|F = 3, m_F = -3\rangle \leftrightarrow |F = 4, m_F = -3\rangle$ transition is probed at half maximum on the blue side of the resonance (0.528 Hz above the resonance center). The fourth fountain cycle, $|F = 3, m_F = +3\rangle$ atoms are selected and the $|F = 3, m_F = +3\rangle \leftrightarrow |F = 4, m_F = +3\rangle$ transition is probed on the blue side of the resonance. This 4180 ms long sequence is repeated so as to implement two interleaved digital servo loops finding the line centers of both the $|F = 3, m_F = -3\rangle \leftrightarrow |F = 4, m_F = -3\rangle$ and the $|F = 3, m_F = +3\rangle \leftrightarrow |F = 4, m_F = +3\rangle$ transitions. With this method, magnetic field fluctuations over timescales longer than 4 s are filtered in the comparison between the two transition frequencies. Every 400 fountain cycles, the above sequence is interrupted and the regular clock transition $|F = 3, m_F = 0\rangle \leftrightarrow |F = 4, m_F = 0\rangle$ is measured for 10 s allowing for an absolute calibration of the local frequency reference with a suitable statistical uncertainty. The overall statistical uncertainty of the experiment is dominated by the short term ($\tau \leq 4$ s) magnetic field fluctuations (Fig. 6).

We have taken data implementing the experimental sequence described above over a period of 21 days starting on march 30, 2005. The complete raw data (no post-treatment) is shown in Fig. 7, each point representing a ≈432 s measurement sequence of $\omega_{+3} + \omega_{-3} - 2\omega_0$ as described above. Figure 8 shows the frequency stability of the last continuous stretch of data (≈10 days). We note the essentially white noise behavior of the data on Fig. 8, indicating that the experimental

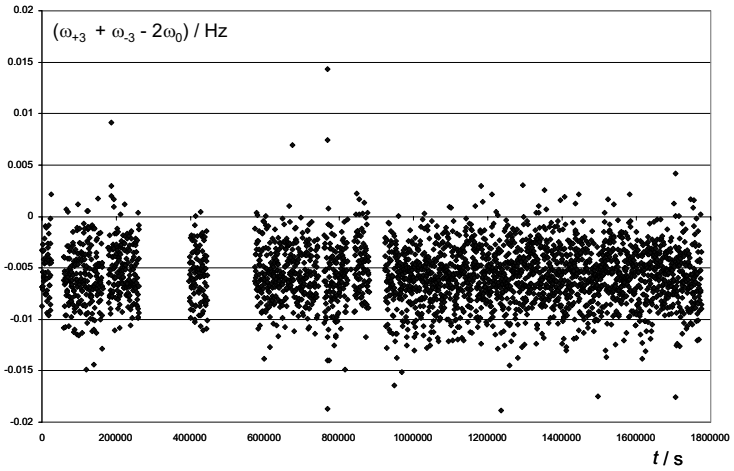


Fig. 7. Raw data of the measurements of $(\omega_{+3} + \omega_{-3} - 2\omega_0)$ spanning ≈21 days

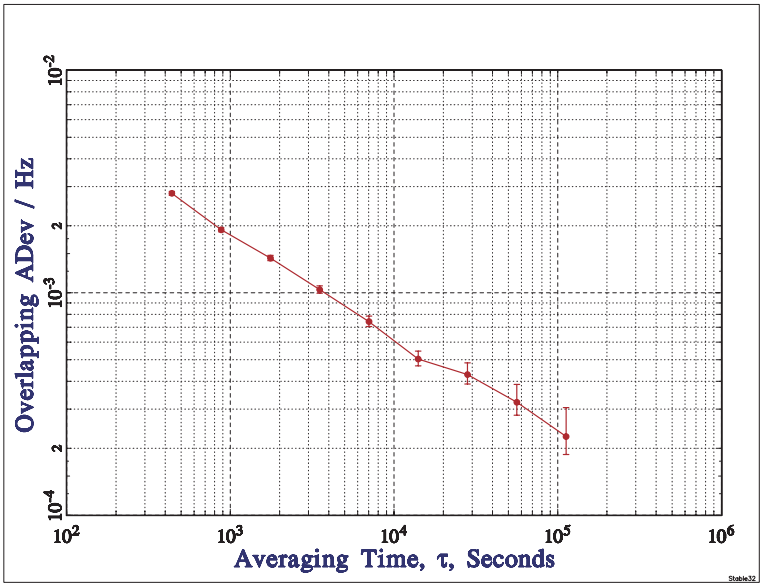


Fig. 8. Frequency stability of the last ≈10 days of the data in Fig. 7

sequence successfully rejects all long term variations of the magnetic field or of other perturbing effects.

According to (16) the frequency of the observable should be the sum of the putative Lorentz violating signal and of the second order Zeeman and other possible systematic corrections. Figure 7 shows a clear offset of the data from

zero, which, using a least squares fit, is found to be (-5.5 ± 0.1) mHz with a very slight linear drift of $(-1.8 \pm 1.0) \times 10^{-7}$ mHz s^{-1}.

For our magnetic field of 202.65 nT the second order Zeeman correction of the $\omega_{+3} + \omega_{-3} - 2\omega_0$ observable is -2.0 mHz. This only partly explains the offset observed in the data. The remaining part is most likely due to the differential influence of the magnetic field on the $m_F = \pm 3$ transitions, resulting from slightly different trajectories of the atoms in the different m_F states and magnetic field inhomogeneities (residual first order Zeeman effect). Such differences in the trajectories could be due to differences in the trapping and/or launching of the atoms, related to the slightly different response of the Zeeman substates to the trapping fields. To check this hypothesis we have looked at the time of flight (TOF) of the atoms as a function of m_F. An offset of $\approx 150\,\mu$s between the $m_F = +3$ and $m_F = -3$ TOF is observed. We are presently studying this effect in more detail (Monte Carlo simulations using the magnetic field map, tests with $m_F = \pm 1$ and $m_F = \pm 2$ states, longer term observation of the TOF difference and its variation, etc.) in order to be able to completely characterize its influence on the offset in Fig. 7, and its variation at sidereal and semi-sidereal frequencies.

In this paper we provide, as a preliminary results, only the values and statistical uncertainties of the coefficients $C_{\omega_\oplus}$, $S_{\omega_\oplus}$, $C_{2\omega_\oplus}$, and $S_{2\omega_\oplus}$ obtained from a model of the form

$$\frac{1}{2\pi}(\omega_{+3} + \omega_{-3} - 2\omega_0) = A\,t + B + C_{\omega_\oplus}\cos(\omega_\oplus t) + S_{\omega_\oplus}\sin(\omega_\oplus t) + C_{2\omega_\oplus}\cos(2\omega_\oplus t) + S_{2\omega_\oplus}\sin(2\omega_\oplus t)\,, \tag{19}$$

and the corresponding order of magnitude limits we expect for the $\tilde{c}^p$ parameters (cf. (17), (18)) of the SME.

Figure 9 shows the amplitudes $A_\omega = \sqrt{C_\omega^2 + S_\omega^2}$ of least squares fits for a range of frequencies including the two frequencies of interest. We note no particularly significant peak at any frequency, and even less so at the frequencies of interest. A least squares fit at those frequencies yields the results shown in Table 9. The correlation coefficients between any two of the four parameters in Table 9 do not exceed 0.07.

From (17), (18) and table I of [31] we deduce orders of magnitude for the limits on the $\tilde{c}^p$ parameters of the SME (see Table 3). We expect to obtain limits on two combinations of the five parameters $\tilde{c}^p_X$, $\tilde{c}^p_Y$, $\tilde{c}^p_Z$, $\tilde{c}^p_Q$, $\tilde{c}^p_-$ at a level of 10^{-25} GeV, and two combinations of the three parameters $\tilde{c}^p_{TX}$, $\tilde{c}^p_{TY}$, $\tilde{c}^p_{TZ}$ at a level of 10^{-19} GeV.

Table 9. Results of the least squares fit of (19) to our complete data. Units are 10^{-5} Hz

$C_{\omega_\oplus}$	$S_{\omega_\oplus}$	$C_{2\omega_\oplus}$	$S_{2\omega_\oplus}$
-5.3 ± 7.3	-10.1 ± 7.2	-3.2 ± 7.2	2.7 ± 7.2

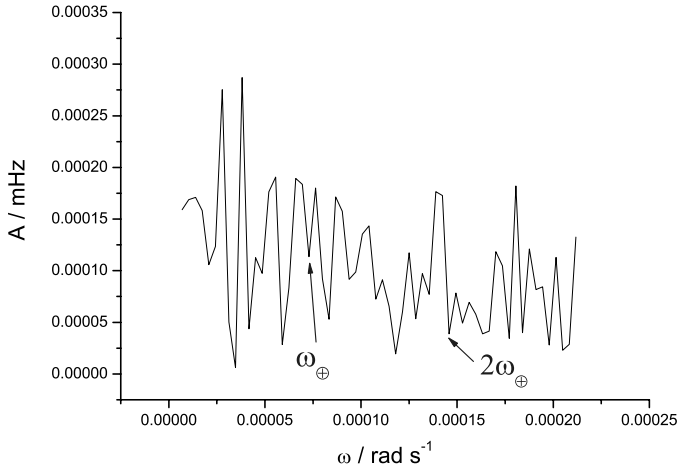


Fig. 9. Fitted Amplitudes A_ω for a range of frequencies around the frequencies of interest (indicated by *arrows*)

In summary, we have carried out an experiment using Zeeman transitions in a cold atom ^{133}Cs fountain clock to test Lorentz invariance in the framework of the matter sector of the SME. In this paper we give a detailed description of the experiment and the theoretical model, we show our data and statistics, and we discuss our still ongoing investigation of systematic effects. Pending the outcome of that investigation and a more detailed theoretical analysis of our experimental results (explicit transformation of $\tilde{c}_q^p$ for our case), we provide only first estimates of the limits that our experiment can set on linear combinations of 8 SME matter sector parameters for the proton. These limits would correspond to first ever measurements of some parameters, and improvements by 11 and 14 orders of magnitude on others. A complete analysis (including systematics) of our experiment with final results for the SME parameters and their uncertainties will be the subject of a near future publication.

5 Conclusion

One hundred years after the publication of Einstein's original paper [1] special relativity, and its fundamental postulate of Lorentz invariance (LLI) are still as "healthy" as in their first years, in spite of theoretical work (unification theories) that hint towards a violation of LLI, and tremendous experimental efforts to find such a violation. Our experiments over the last years have provided some of the most stringent tests of LLI [6–9], but have nonetheless only joined the growing number of experiments in scientific history that measure zero deviation from

LLI, albeit with an ever decreasing uncertainty. In spite of that, experimental tests of LLI are continuing along two lines: decrease of the uncertainties (see for example the contributions on rotating Michelson-Morley experiments in this volume) on one hand, and new types of experiments, e.g. the atomic clock test reported here, on the other.

In this paper we have presented a review of our recent Michelson-Morley and Kennedy-Thorndike experiment (Sect. 3), and reported first results of our ongoing experiment that tests Lorentz invariance in the matter sector using a cold Cs atomic fountain clock (Sect. 4). We have briefly described the two theoretical frameworks used to model and analyze our experiments (the Robertson-Mansouri-Sexl (RMS) framework and the standard model extension (SME)), and derived experimental limits on a number of parameters of those frameworks. When compared to other experiments those limits are the most stringent at present for several parameters (see tables 1, 2, 3).

The next generation of Michelson-Morley experiments are based on similar technology as our experiment (Sect. 3) or the equivalent approach at optical frequencies [25], but take advantage of active rotation of the experiment (see the corresponding contributions in this volume). Rotation of the experiment (typically at about 0.1 Hz) allows much faster data integration and places the signal modulation frequency close to the optimum where resonators are the most stable. It is expected that such experiments will lead to order(s) of magnitude improvements on orientation dependent parameters in the theoretical frameworks, but they present no advantage for only velocity dependent parameters. For example, in the RMS rotating experiments are likely to provide new, more stringent limits for the Michelson-Morley parameter ($P_{MM} = 1/2 - \beta_{\rm MS} + \delta_{\rm MS}$) but no improvements on the Kennedy-Thorndike one ($P_{KT} = \beta_{\rm MS} - \alpha_{\rm MS} - 1$). So we expect our (and other) present limits on P_{MM} to be significantly improved, but we see no obvious way of improving on our present limit on P_{KT} in the near future.

Several improvements of our clock test of LLI in the SME matter sector (Sect. 4) are possible. For example, using the unique capability of our double fountain (FO2) to run on both, Cs and Rb, we expect to be able to use Cs as the SME sensitive species and Rb (which is less sensitive to the SME [31]) as the magnetic field probe. In that way we should be able to perform magnetic field independent measurements that could improve on our present results, and allow access to other SME parameters that we are insensitive to with our present set up. Also, rotation of the experiment could provide a method for faster modulation of the signal but is unpractical in an Earth bound laboratory. However, space missions with onboard atomic clocks are well suited for such a test. In particular the European ACES (Atomic Clock Ensemble in Space) mission [36], scheduled for flight on the international space station (ISS) in 2009, seems very promising in this respect. It will include a laser cooled Cs clock (PHARAO) with expected performance at least equivalent to our FO2, but with the orientation of its quantization field axis modulated at a 90 min period (ISS orbital

period) rather than 24 hr as in our case. This should allow for much faster data integration and significant improvement on the limits presented here.

References

1. Einstein A., Annalen der Physik, **17**, 891-921, (1905).
2. Kostelecky V.A., Samuel S., Phys. Rev. **D39**, 683, (1989).
3. Damour T., gr-qc/9711060 (1997).
4. Gambini R., Pullin J., Phys. Rev. **D59**, 124021, (1999).
5. Will C.M., *Theory and Experiment in Gravitational Physics, revised edition*, Cambridge U. Press, (1993).
6. Wolf P. and Petit G., Phys. Rev. **A56**, 6, 4405, (1997).
7. Wolf P. et al., Phys. Rev. Lett. **90**, 6, 060402, (2003).
8. Wolf P., et al., Gen. Rel. and Grav. **36**, 10, 2351–2372, (2004).
9. Wolf P., et al., Phys. Rev. **D70**, 051902(R), Rapid Communication, (2004).
10. Robertson H.P., Rev. Mod. Phys. **21**, 378 (1949).
11. Mansouri R. and Sexl R.U., Gen. Rel. Grav. **8**, 497, 515, 809, (1977).
12. Colladay D., Kostelecký V.A., Phys. Rev. **D55**, 6760, (1997); Colladay D., Kostelecký V.A., Phys. Rev. **D58**, 116002, (1998); Kostelecký V.A., Phys. Rev. **D69**, 105009, (2004).
13. Tobar M.E., et al., Phys. Rev. **D71**, 025004, (2005).
14. Saathoff G., et al., Phys. Rev. Lett. **91**, 19, 190403, (2003).
15. Lightman A.P. and Lee D.L., Phys. Rev. **D8**, 2, 364, (1973).
16. Blanchet L., Phys. Rev. Lett. **69**, 4, 559, (1992).
17. Ni W.-T., Phys. Rev. Lett. **38**, 301, (1977).
18. Damour T. and Polyakov A.M., Nucl. Phys. **B423**, 532, (1994).
19. Kostelecky A.V. and Mewes M., Phys. Rev. **D66**, 056005, (2002).
20. Fixsen D.J. et al., Phys. Rev. Lett. **50**, 620, (1983).
21. Lubin et al., Phys. Rev. Lett. **50**, 616, (1983).
22. Michelson A.A. and Morley E.W., Am. J. Sci., **34**, 333, (1887).
23. Brillet A. and Hall J.L., Phys. Rev. Lett. **42**, 9, 549, (1979).
24. Kennedy R.J. and Thorndike E.M., Phys. Rev. **B42**, 400, (1932).
25. Müller H. et al., Phys. Rev. Lett. **91**, 2, 020401, (2003).
26. Ives H.E. and Stilwell G.R., J. Opt. Soc. Am. **28**, 215 (1938).
27. Braxmaier C. et al., Phys. Rev. Lett. **88**, 1, 010401, (2002).
28. Hils D. and Hall J.L., Phys. Rev. Lett., **64**, 15, 1697, (1990).
29. Kostelecky V.A., Phys. Rev. **D69**, 105009, (2004).
30. Kostelecky and Lane, Phys. Rev. **D60**, 116010, (1999).
31. Bluhm R., Kostelecky A.V., et al., Phys. Rev. **D68**, 125008, (2003).
32. Bize S. et al., Proc. 6th Symp. on Freq. Standards and Metrology, World Scientific, (2002).
33. We assume here that local position invariance is sufficiently verified so that the variation of the maser frequency due to the diurnal variation of the local gravitational potential is negligible. Indeed the results of [34] imply that such variations should not exceed 2 parts in 10^{-17} which is significantly below our noise level.
34. Bauch A., Weyers S., Phys. Rev. **D65**, 081101, (2002).
35. Tobar M.E. et al., Phys. Lett. **A300**, 33, (2002).
36. Salomon C. et al., C.R. Acad. Sci. Paris, **2**, 4, 1313, (2001).
37. Müller H. et al., Phys. Rev. **D67**, 056006, (2003).

38. Parker T.E., IEEE Trans. UFFC, **46**, 745, (1999).
39. Chang S., Ph.D. thesis, Univ. of West. Aus., (2000).
40. Colladay D. and Kostelecky V.A., Phys. Rev. **D55**, 6760, (1997); **D58**, 116002, (1998).
41. Cané F. et al., Phys. Rev. Lett. **93**, 230801, (2004).
42. Müller H., Phys. Rev. **D71**, 045004, (2005).
43. Lane C.D., arXiv:hep-ph/0505130.
44. Phillips D.F. et al., Phys. Rev. **D63**, 111101(R), (2001); Humphrey M.A., arXiv:physics/0103068; Phys. Rev. **A62**, 063405, (2000).
45. Bear D. et al., Phys. Rev. Lett. **85**, 5038, (2000).
46. Hou L.-S., Ni W.-T., Li Y.-C.M., Phys. Rev. Lett. **90**, 201101, (2003); Bluhm R. and Kostelecky V.A., Phys. Rev. Lett. **84**, 1381, (2000).
47. Bize S., et al., J. Phys. **B38**, S449-S468, (2005).
48. Marion H., et al., Phys. Rev. Lett. **90**, 150801, (2003).
49. Vanier J., Audoin C., *The Qunatum Physics of Atomic Frequency Standards*, Adam Hilger, (1989).

Experimental Test of Time Dilation by Laser Spectroscopy on Fast Ion Beams

G. Saathoff[1], G. Huber[2], S. Karpuk[2], C. Novotny[2], S. Reinhardt[1], D. Schwalm[1], A. Wolf[1] and G. Gwinner[3]

[1] Max-Planck-Institut für Kernphysik, 69029 Heidelberg, Germany
guido.saathoff@mpi-hd.mpg.de
[2] Institut für Physik, Universität Mainz, 55099 Mainz, Germany
[3] Department of Physics and Astronomy, University of Manitoba, Winnipeg, MB R3T 2N2, Canada

1 Introduction

The Ives-Stilwell experiment – measuring the time dilation relation $\gamma_{\mathrm{SR}} = (1-\beta^2)^{-1/2}$ – is one of the three classic experiments to test Special Relativity. Together with the interferometric tests of the velocity-independence as well as the isotropy of the speed of light, governed by the Michelson-Morley and the Kennedy-Thorndike experiment, respectively, the Ives-Stilwell experiment entirely establishes Special Relativity on an experimental basis and replaces Einsteins postulates [1]. While the interferometric experiments are 'null-experiments' looking for deviations of the constancy of the speed of light, the time dilation test provides a positive measurement based on a Lorentz boost. It was Einstein who proposed already in 1907 to look for the time dilation effect by observing the Doppler-shifted wavelength of light emitted by excited fast atoms perpendicularly to the motion. In this direction, the classical Doppler effect vanishes leaving pure time dilation. However, it turned out that this scheme is difficult to implement as already small deviations of the observation angle from 90 degree would cause frequency shifts due to the classical Doppler effect which varies linearly around $\pi/2$. It took another 31 years until Ives and Stilwell performed the first measurement. Contrary to Einsteins idea they observed the Doppler shifts not perpendicularly but in forward and in backward direction of the atomic motion. This scheme provides different advantages. First, the classical Doppler shift vanishes in this scheme because it is of equal magnitude but opposite in sign in both directions of observation. Secondly, the measurement is less sensitive to small misalignments as the classical Doppler shift varies only quadratically around 0 and π. And finally, as two frequencies are measured, both the time dilation factor as well as the atoms' velocity can be extracted to an accuracy given by the frequency uncertainty. When observing perpendicularly, the velocity has to be determined separately to provide a test of the time dilation relation. In the next section the principle of the Ives-Stilwell experiment

G. Saathoff et al.: *Experimental Test of Time Dilation by Laser Spectroscopy on Fast Ion Beams*, Lect. Notes Phys. **702**, 479–492 (2006)
DOI 10.1007/3-540-34523-X_17
© Springer-Verlag Berlin Heidelberg 2006

will be presented. Section 3 deals with the implementation of this experiment at a heavy ion storage ring. This will include the role of the ions used and a brief discussion of the basic properties of the TSR storage ring in Heidelberg as relevant for the experiment. Moreover the experimental setup is described and two types of spectroscopy are investigated. Finally the current status is given together with future prospects.

2 Principle of the Ives Stilwell Experiment

In the original Ives-Stilwell experiment [2] excited hydrogen atoms in canal rays are used as clocks moving at a velocity $\beta = v/c = 0.005$ with respect to the laboratory frame S_{lab}. The Doppler-shifted frequencies ν_{p} and ν_{a} of the Hβ line are measured in parallel ($\vartheta_{\mathrm{p}} = 0$) and antiparallel ($\vartheta_{\mathrm{a}} = \pi$) direction with respect to $\boldsymbol{\beta}$, using a grating spectrometer. Within Special Relativity the respective Doppler-shifts are given by the relativistic Doppler-formula, $\nu_0 = \gamma_{\mathrm{SR}}(\beta^2) \cdot (1 - \beta \cos \vartheta_{\mathrm{p,a}})\nu_{\mathrm{p,a}}$, which is a direct consequence of time dilation. ν_0 is the frequency of the transition in the system S_{rest}, in which the hydrogen atom is at rest. Multiplication of these equations yields the velocity-independent relation $\nu_{\mathrm{p}}\nu_{\mathrm{a}} = \nu_0^2$, since the Special Relativity time dilation factor $\gamma_{\mathrm{SR}}^2(\beta^2)$ obeys $\gamma_{\mathrm{SR}}^2(\beta^2) \cdot (1 - \beta^2) = 1$.

Possible deviations of time dilation from γ_{SR} can be quantified using the kinematical test theory by Robertson [1], which was later modified by Mansouri and Sexl [3]. Allowing the existence of a hypothetical preferred frame $\Sigma(T, \mathbf{X})$, in which the speed of light c_0 is assumed to be isotropic, they consider generalized Lorentz transformations between Σ and a frame $S(t, \mathbf{x})$ moving relative to Σ at a velocity V along the X axis. Using Einstein synchronization, these transformations read

$$T = \Gamma \left(\frac{t}{\hat{a}} + \frac{Vx}{\hat{b}\, c_0^2} \right); \tag{1}$$

$$X = \Gamma \left(\frac{x}{\hat{b}} + \frac{Vt}{\hat{a}} \right);\ Y = \frac{y}{\hat{d}};\ Z = \frac{z}{\hat{d}}$$

with $\Gamma = (1 - V^2/c_0^2)^{-\frac{1}{2}}$. Due to the abolition of the relativity principle these transformations are in general only valid with respect to Σ and not between two arbitrary, constantly moving reference frames. The velocity-dependent test functions $\hat{a}(V^2)$, $\hat{b}(V^2)$, and $\hat{d}(V^2)$ modify time dilation as well as Lorentz contraction in longitudinal and transverse direction. They reduce to $\hat{a}(V^2) = \hat{b}(V^2) = \hat{d}(V^2) = 1$ in case SR holds. For most experiments these functions can be reduced to three velocity-independent parameters $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\delta}$ in the low-velocity limit: $\hat{a}(V^2) = [1 + \hat{\alpha}V^2/c_0^2 + \mathcal{O}(c_0^{-4})]$, $\hat{b}(V^2) = [1 + \hat{\beta}V^2/c_0^2 + \mathcal{O}(c_0^{-4})]$, and $\hat{d}(V^2) = [1 + \hat{\delta}V^2/c_0^2 + \mathcal{O}(c_0^{-4})]$.

In this test theory, the speed of light $c(\theta, V)$ in the moving frame S,

$$\frac{c(\theta,V)}{c_0} = 1 + (\hat{\beta} - \hat{\delta})\frac{V^2}{c_0}\sin^2(\theta) + (\hat{\alpha} - \hat{\beta})\frac{V^2}{c_0} \tag{2}$$

is in general not constant, but depends on the angle θ between the direction of $c(\theta, V)$ and the motion of the moving frame S as well as on the velocity V between Σ and S. The Michelson-Morley experiment [4] is sensitive to an anisotropy of the speed of light thus determining the parameter combination $|\hat{\beta} - \hat{\delta}|$ and the Kennedy-Thorndike experiment [5] tests the velocity-dependence of c described by $|\hat{\alpha} - \hat{\beta}|$. The Ives-Stilwell experiment independently measures the parameter $\hat{\alpha}$ that describes time dilation.

A detailed analysis of the latter within the Mansouri–Sexl test theory has been carried out in [6] and shows that a non-vanishing test parameter $\hat{\alpha}$ would modify the outcome of the Ives-Stilwell experiment as

$$\frac{\nu_\mathrm{p}\nu_\mathrm{a}}{\nu_0^2} = 1 + 2\,\hat{\alpha}\,(\beta^2 + 2\,\boldsymbol{\beta}_\mathrm{lab}\cdot\boldsymbol{\beta}) + \mathcal{O}(c^{-4}), \tag{3}$$

with $\boldsymbol{\beta}_\mathrm{lab} = \mathbf{V}/c_0$. The β^2 term allows to determine $\hat{\alpha}$ absolutely without having to rely on the precise knowledge of β_lab (at least as long as β is larger than β_lab), while the $2\boldsymbol{\beta}_\mathrm{lab}\boldsymbol{\beta}$ term gives access to $\hat{\alpha}$ via sidereal modulations. To derive $\hat{\alpha}$ from the latter, $\boldsymbol{\beta}_\mathrm{lab}$ is usually regarded as the velocity of the earth with respect to the cosmic microwave background rest frame, which provides a plausible but not necessary candidate for a preferred reference frame.

The original IS experiment provided an upper bound of $|\hat{\alpha}| < 1 \times 10^{-2}$ from the β^2 term. Later, significant improvements have been achieved using laser techniques instead of conventional spectrometers; two-photon spectroscopy on a $\beta = 0.0036$ neon atomic beam has set an absolute bound of $|\hat{\alpha}| < 2.3 \times 10^{-6}$ [7] and, considering the 3 K cosmic microwave background rest frame as the preferred one ($\beta_\mathrm{lab}c \approx 350\,\mathrm{km/s}$), even $|\hat{\alpha}| < 1.4 \times 10^{-6}$ [8] from the limit on sidereal variations. A different way of testing time dilation was carried out by P. Wolf [9]. In this work, GPS data was analyzed leading to $|\hat{\alpha}| < 1 \times 10^{-6}$ from sidereal variations.

3 Ives-Stilwell Experiment at Storage Rings

A further improvement of this limit has been made possible by the development of heavy-ion storage rings like the TSR in Heidelberg [10], which provide high-quality particle beams at a considerably higher velocity. In this work, two different approaches of Doppler-free spectroscopy have been investigated, saturation spectroscopy on a closed two-level system and Λ-spectroscopy on a three-level system.

3.1 $^7Li^+$ Ions as Moving Clocks

For the Doppler shift measurement, the ions used as moving clocks are required to have a transition in the optical region, which is accessible by lasers from

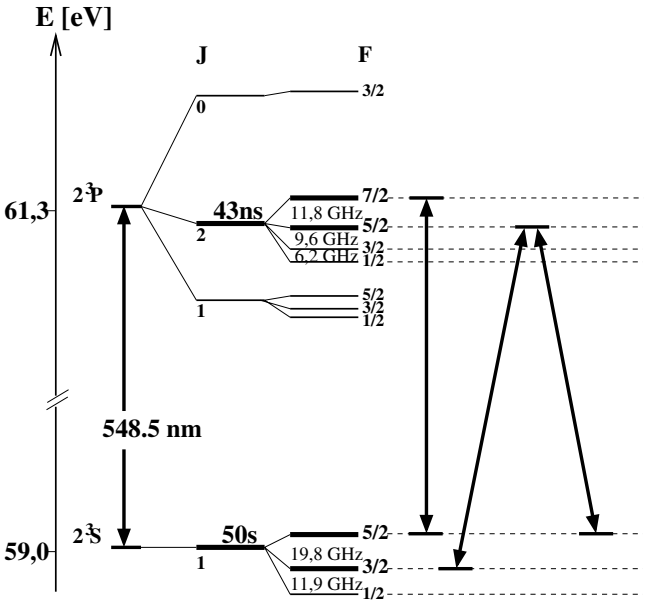


Fig. 1. Level-scheme of the triplet system in $^7Li^+$, which is ortho-heliumlike. It contains a closed Λ-system, which is used for Doppler-free Λ-spectroscopy, and a closed two-level system suitable for saturation spectroscopy

forward and backward direction even for large Doppler shifts. Furthermore, this transition should be narrow to allow for high resolution spectroscopy and its rest-frame frequency must be known to high accuracy. These requirements restrict the experiment to the heliumlike $^7Li^+$ ion. It has a strong optical transition $2s\ ^3S_1 \rightarrow 2p\ ^3P_2$ at 548.5 nm in the metastable triplet system shown in Fig. 1. The lower level of this transition lies 59 eV above the singlet ground state. It has a radiative lifetime of 50 seconds in vacuum, which is, however, decreased to about 15 seconds mainly by stripping due to collisions with the rest gas when stored in the TSR at a typical rest gas pressure of 5×10^{-11} mbar. The lifetime of the upper level is 43 ns [11] corresponding to a natural linewidth of the transition of 3.8 MHz. Both Doppler-shifted transition frequencies ν_a and ν_p lie in the optical region, even for very high ion velocities of $\beta = 0.4$ which will be accessible at the ESR storage ring of the Gesellschaft für Schwerionenforschung in Darmstadt. The $^7Li^+$ ion has a nuclear spin of 3/2 leading to the hyperfine structure shown in Fig. 1. The splittings between these hyperfine structure components are of the order of 10 to 20 GHz in the $^3S_1$ and $^3P_2$ levels, which is larger than the Doppler width caused by the velocity distribution of the ion beam. Therefore, the transitions $2s\ ^3S_1(F = 5/2) \rightarrow 2p\ ^3P_2(F = 5/2)$ and $2s\ ^3S_1(F = 3/2) \rightarrow 2p\ ^3P_2(F = 5/2)$ form a closed three-level Λ-system, and the $2s\ ^3S_1(F = 5/2) \rightarrow 2p\ ^3P_2(F = 7/2)$ transition provides a closed two-level system.

Both level schemes have been used for Doppler shift measurements. The Λ-system is well suited for optical-optical double resonance spectroscopy as

described in Sect. 3.4, and the two-level system makes saturation spectroscopy possible. This is described in Sect. 3.5.

3.2 The Heidelberg Test Storage Ring (TSR)

The TSR basically consists of 4 pairs of 2 45-degree dipole magnets, which bend the ions on a closed orbit with a circumference of about 55 m. 20 magnetic quadrupole lenses accomplish the focusing of the beam. The velocity at which Li ions can be stored in the TSR is limited to $\beta = 0.064$ by the maximum magnetic rigidity of the TSR dipoles of 1.5 Tm. It leads to Doppler shifts of the 548 nm transition to 514 nm in the parallel and 585 nm in the antiparallel direction, respectively. The ion beam at the corresponding energy of 13.3 MeV is provided by a tandem Van-de-Graaff accelerator. Starting with negative Li-ions, about 10% of the ions emerge in the metastable $^3S_1$ state of Li^+ from the gas stripper at the terminal of the accelerator and typically 10^8 ions are injected into the TSR. Shortly after finishing the injection process, typical parameters of the ion beam in the experimental section are a diameter of 2 cm, a divergence of 2 mrad and a longitudinal energy spread of $\sigma_p/p = 2 \times 10^{-3}$, making a precise frequency measurement impossible. To improve its brilliance, the ion beam is subjected to electron cooling during the whole measurement. The Coulomb interaction of the ions with a cold electron beam, co-moving at the same speed over a length of 1.5 m, leads to a friction force in both the longitudinal and the transverse direction. As a result, not only the longitudinal velocity spread is strongly decreased, but also the beam diameter as well as the divergence shrink by a factor of 40. At equilibrium, which is reached after about 7 s of cooling, the ion beam in the ion-laser interaction region (see Fig. 2) has a σ-width of $\approx$250 µm, a σ-divergence of $\approx$50 µrad, and a longitudinal momentum spread of $\sigma_p/p = 3.5 \times 10^{-5}$. The latter leads to a Doppler-width of the transition of about 2.5 GHz (FWHM), which is much smaller than the hyperfine splittings of the involved levels [12]. Moreover due to the friction force, the ions adapt the velocity of the electrons, which allows for an accurate control and stabilization of the ion velocity. To compensate for remaining slow drifts of the mean ion velocity, which are due to small variations of the acceleration voltage of the electrons in the electron cooler, the ion beam is moderately bunched by applying a sinusoidal rf-voltage at the 3rd harmonic of the average revolution frequency. The ions are confined in a resulting co-moving pseudopotential, which fixes and controls the revolution frequency and, consequently, the ion velocity.

3.3 The Experimental Setup

Figure 2 shows the experimental setup at the TSR. The blue-shifted light at 514 nm, overlapped parallel to the ion beam, is provided by a fixed-frequency Ar^+ laser. It is first locked to a high-finesse Fabry-Perot cavity in order to compensate for short-term fluctuations. Long-term stabilization and frequency calibration is accomplished by locking the laser to a well-known hyperfine structure line in

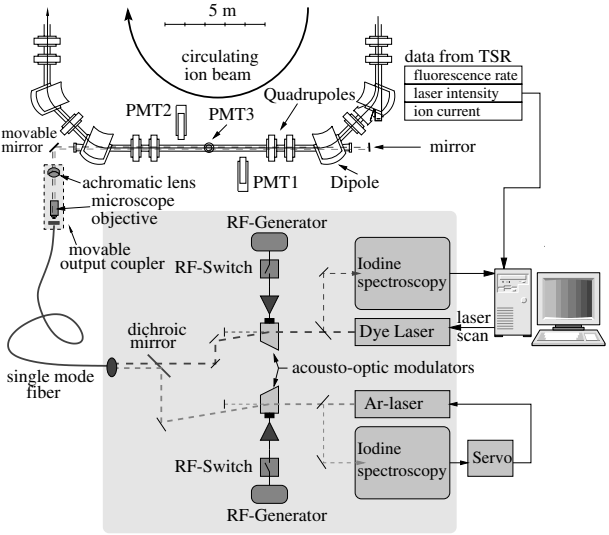


Fig. 2. Setup of the experiment at the ion storage ring TSR (from [14])

$^{127}I_2$ using saturation spectroscopy. Containing all systematic errors in the laser stabilization, the uncertainty of the Ar^+ laser frequency amounts to 61 kHz.

The red-shifted 585 nm light for the parallel excitation is provided by a tunable dye laser. It has a linewidth of below 1 MHz and is scanned across the $^7Li^+$ resonance, typically over a range of 200 MHz. In order to determine the frequency of the scanning laser, saturation spectroscopy is performed on a suitable $^{127}I_2$ hyperfine structure line simultaneously to the Li^+ spectroscopy.

To minimize the error in the frequency calibration, the dye as well as the Ar^+ laser beam going into the TSR are passed through acousto-optic frequency shifters in order to shift the Lamb dip very close to the frequency of the I_2 line, used for the calibration of the dye laser. The acousto-optic modulators are additionally used to chop the laser beams going to the TSR at frequencies of several kHz using rf-switches. The two laser beams are then merged with a dichroic mirror and guided to the TSR by a single-mode polarization-maintaining fiber. The bichromatic beam is then expanded in an achromatic telescope, directed through the experimental section of the TSR, and retro-reflected by a plane mirror. Both laser beams are linearly polarized in the same direction and their intensities are matched to balance laser forces on the ion beam. The focus of the bichromatic beam is placed onto the mirror in order to mode-match the incoming and the reflected beam. Three photomultipliers at different positions along the interaction region detect the fluorescence from the Lithium ions.

3.4 Optical-Optical Double Resonance Spectroscopy

To derive time dilation from observations of the Doppler shifts in parallel and antiparallel direction, the classical part of the Doppler effect has to be overcome.

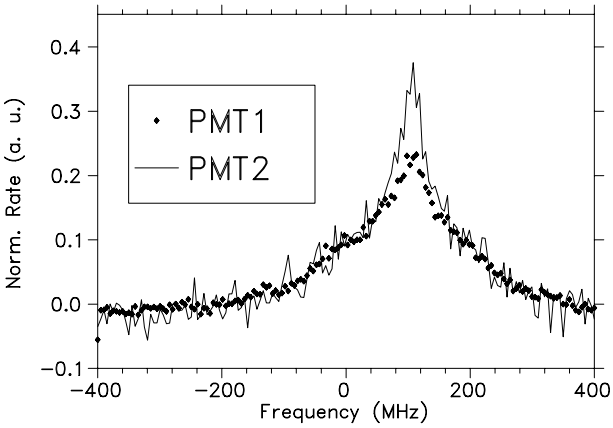


Fig. 3. Λ-resonance taken with to photomultipliers at different positions along the interaction beam line. The signal from PMT2 is scaled to match the PMT1 signal at the wings. The additional fluorescence seen in PMT2 stems from ions interacting with both legs of the Λ within one passage through the interaction region

The first method that was applied to do a first order Doppler free measurement was optical-optical double resonance spectroscopy on the Λ-system shown in Fig. 1. Each laser beam is tuned into resonance with one of the legs of the Λ. Only when both lasers are on resonance with an ion, strong fluorescence is observed, as a single resonant laser quickly pumps the ion into the other, "dark" ground state. Due to the narrow bandwidths of the lasers they can only be simultaneously in resonance with a narrow velocity class, yielding (to first order) a Doppler-free "Λ-resonance". In resonance condition (3), ν_0 has to be replaced by the product $\nu_1\nu_2$ of the transition frequencies involved in the Λ. The first storage ring test of time dilation has employed this method in 1994 and has set the thitherto best absolute bound of $|\hat{\alpha}| < 8 \times 10^{-7}$ [12]. The result was limited by the large observed linewidth of the Λ-resonance of almost 30 MHz[4], compared to a natural width of 3.8 MHz. To understand this broadening, a more detailed investigation of the double resonance spectroscopy has been performed. Figure 3 shows a measurement of the Λ-resonance with two photomultipliers located at different positions along the beam line (PMT1 and PMT2 in Fig. 2). After subtraction of the background from both spectra, the signal of PMT2 could be scaled to perfectly match the PMT1-signal in the wings, but revealing additional fluorescence in the peak center. We interpret the broad background fluorescence, showing the same shape in both photomultiplier signals, as stemming from ions that have been pumped dark by any of the lasers before, but come back into resonance with the other laser because of velocity changes they experience during several to many roundtrips in the storage ring. The extra fluorescence seen by PMT2 stems from ions interacting with both lasers during one passage through

[4] Note that the observed width of the resonance is twice the actual linewidth as only one of the lasers is scanned.

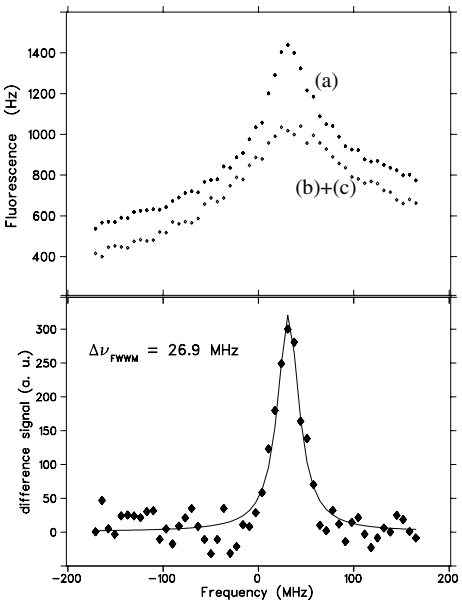


Fig. 4. Λ-resonance taken with a laser chopping frequency of 200 kHz. *Upper panel*: Spectrum with both lasers on simultaneously (trace a) and sum of the spectra with the lasers on separately (trace b+c). *Lower panel*: Difference signal fitted with a Lorentzian

the interaction region. The reason, why PMT2 sees more of this "true" Λ-signal is that it is located further downstream, so that most of the dark-pumped ions will interact with the resonant laser shortly after entering the interaction region. This interpretation is further confirmed by a quasi-simultaneous measurement of 3 spectra taken with different laser configurations. The lasers are switched between these configurations with the AOMs at a frequency in the kHz to 100 kHz range. One spectrum (a), taken with both lasers on simultaneously, provides the complete Λ-signal. Two spectra (b) and (c) are taken with the lasers switched on separately and contain the fluorescence stemming only from ions that change their velocity slower than the chopping, but lack the "true" Λ-signal as well as the fluorescence from the fast velocity changes. Figure 4 (upper panel) shows spectrum (a) as well as the sum of (b) and (c), which is drawn with an offset for clarity. The lower panel shows the difference, which basically contains the true signal and can well be fitted with a Lorentzian. Moreover Fig. 5 shows that, as we increase the chopping frequency, the difference signal becomes narrower, because faster velocity-changes are detected in the sum spectrum and therefore subtracted.

Due to the necessity to subtract most of the signal, the remaining undisturbed resonance shows to be rather weak limiting the accuracy of a frequency measurement.

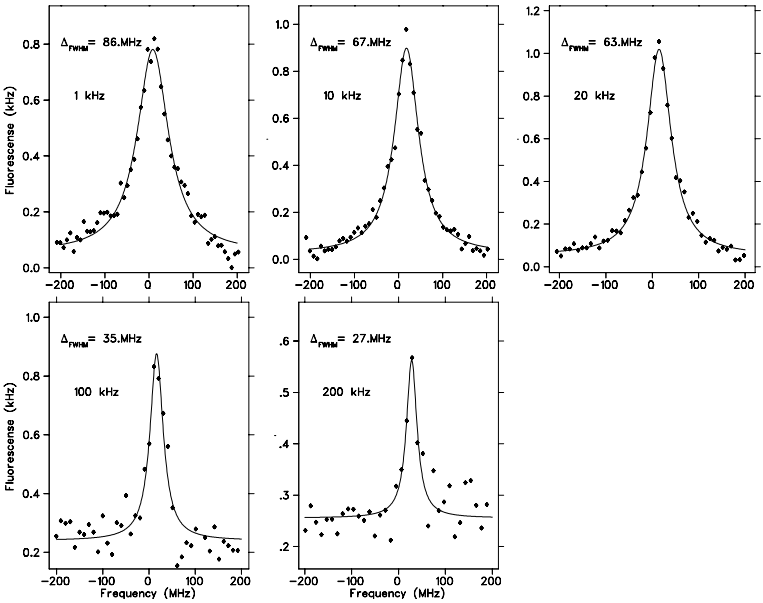


Fig. 5. Difference spectra of Λ-resonances taken at different chopping frequencies. With increasing frequency, fluorescence due to faster velocity-changes are subtracted from the total Λ-signal

3.5 Saturation Spectroscopy

In order to circumvent the problems associated with the Λ-spectroscopy, we instead applied saturation spectroscopy on the $^7\mathrm{Li}^+$ $2^3S_1(F = 5/2) \longrightarrow 2^3P_2(F = 7/2)$ two-level transition. The laser intensities are chosen equal and sufficiently high to saturate the transition leading to a nonlinear dependence of the fluorescence rate on the intensity. The narrow-bandwidth lasers are in general talking to ions of different velocity classes, so that their fluorescence yields simply add up. Only when both lasers are on resonance with the same velocity class, the ions are saturated and the corresponding total fluorescence yield is less, leading to a narrow Lamb dip. The main difference of the saturation spectroscopy on ion beams in storage rings as compared to gas cells is the particle density. Due to the high laser intensities required to saturate the transition, correspondingly strong laser forces on the ions are unavoidable. While the resulting changes in the velocity distribution, which could influence the shape of the fluorescence background, are quickly leveled out in gas cells due to the high scattering rate, velocity changes in the ion beam can prevail considerably longer and thus distort the background and the Lamb dip. Therefore, we chose the laser intensities to be equal in order to balance these forces rather than applying one pump and one probe beam like commonly done in gas cells. Moreover due to the low absorption, we have to detect the fluorescence as this is much more sensitive. In contrast to the Λ-type measurement, saturation spectroscopy requires the interaction of

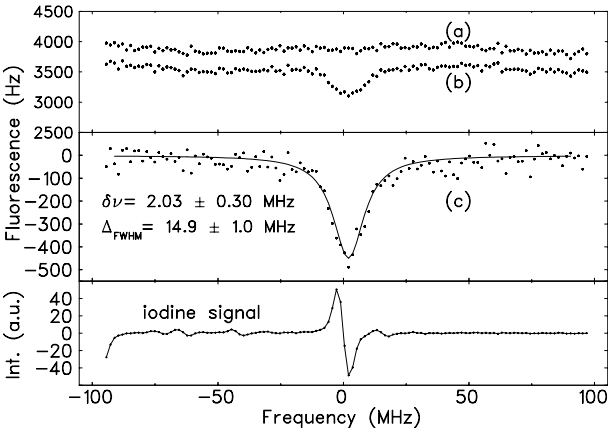


Fig. 6. Fluorescence signal observed with PMT3 for a multiple scan of the Lamb dip. Spectrum (a) is plotted with an offset for clarity. The zero of the frequency scale corresponds to the position of the iodine reference line. (from [14])

both lasers with an individual ion within the time-scale of the spontaneous decay of the excited state, which is 43 ns. Velocity-changing processes occur on longer time-scales and, hence, do not influence the width of the Lamb dip.

However, the Lamb dip lies on a large fluorescence background which is decreasing during a laser scan due to the exponential decay of the metastable ions. We therefore decouple the laser scan cycle from the ion injection cycle by taking only a non-integer part of the spectrum (46 of the 200 data points of one laser scan) at each injection. The laser scan is stopped and a fresh ion beam is injected and prepared in the storage ring, on which the next 46 data points are taken. Figure 6 (trace b) shows a run containing 82 laser scans where the exponential decay is largely averaged out. To account for the remaining fluctuations in the fluorescence signal, which are mainly due to varying ion numbers from different injections, the Doppler background is measured quasi-simultaneously with the Lamb dip using the same chopping scheme as for the Λ spectroscopy. Trace (b) shows the spectrum taken with both lasers on. It contains the Doppler background together with the Lamb dip. Trace (a) is the sum of the spectra recorded with the lasers on separately and it contains the Doppler background only. The difference of (a) and (b) (trace c) shows the pure Lamb dip.

Several systematic error sources have been studied. First we investigated the influence of the laser forces on the spectrum by varying the laser intensity. Fig. 7 shows a slight dependence of the Lamb dip position $\delta\nu$ on the total laser intensity. We attribute this to local changes in the velocity distribution caused by the laser forces that occur faster than our background subtraction time scale which is set by the chopping frequency of 5 kHz. To account for the residual effect, we extrapolate the resonance frequency to zero intensity by fitting $\delta\nu = \delta\nu_{\mathrm{L}} + mI^{\kappa}$ with m, κ and $\delta\nu_{\mathrm{L}}$ as fit parameters, yielding an almost linear dependence ($\kappa = 0.93$) with a frequency offset at intensity zero of $\delta\nu_{\mathrm{L}} =$

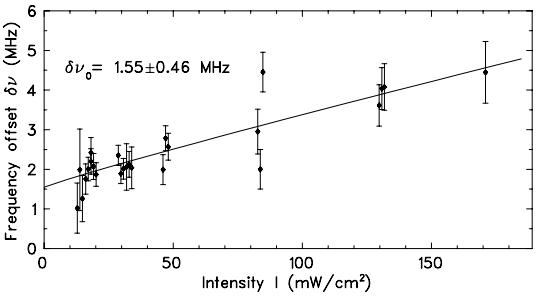


Fig. 7. Frequency offset of the Lamb dip from the I_2 reference line as a function of the total laser intensity $I = I_a + I_p$ with the respective parallel and antiparallel laser intensities I_p and I_a kept equal. The laser switching frequency is 5 kHz (from [14])

1550±460 kHz relative to the iodine line. For future experiments, we will increase the chopping frequency to 50 kHz, which has recently proven to largely remove the intensity dependence. Figure 8 shows a comparison of the two measurements done wit 5 kHz and 50 kHz, respectively. For fast chopping, a linear as well as a constant fit to the data yields essentially the same result and the fit uncertainty is improved to the 100 kHz range. Moreover, the extrapolated Lamb dip frequencies from measurements taken at the different frequencies coincide well within the uncertainty and confirm the fit result in Fig. 7.

Magnetic strayfields of about 0.5 Gauss near the photomultiplier used for the frequency measurement (PMT3 in Fig. 2) cause Zeeman shifts of the magnetic sublevels of the order of several hundred kHz. In order to prevent a corresponding frequency shift of the transition that may be even enhanced by optical pumping, we apply linearly polarized light only causing a slight broadening but not a significant net shift of the resonance.

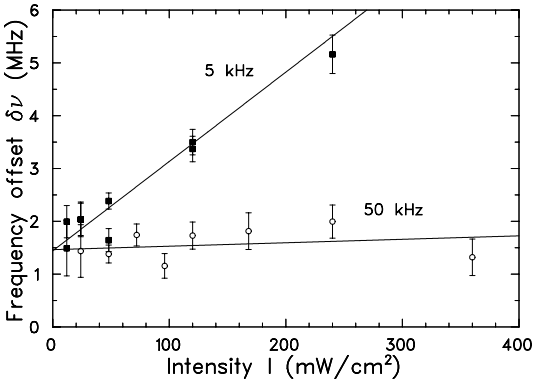


Fig. 8. Comparison of the intensity dependence of the Lamb dip position for two different laser chopping frequencies

Misalignments of the laser-ion angles ϑ_p and ϑ_a from 0 and π cause frequency shifts. From Lamb dips recorded at different angles a shift as expected when assuming plane waves was confirmed [13]. We are able to optimize and monitor the overlap of the lasers with the ion beam to below 70 μrad by moving the laser beams transversally to the ion beam and adjusting the angle for simultaneous occurrence of the fluorescence at three photomultipliers located at different positions along the beam line. The corresponding frequency error amounts to below 10 kHz. The misalignment of the retro-reflected Ar laser can also be limited to 70 μrad causing a frequency shift of below 40 kHz. The remaining divergence of the electron-cooled ion beam is 50 μrad and causes a frequency shift of less than 10 kHz, whereas an uncooled beam would shift the Lamb dip by 3 MHz.

A further influence can be caused by the Gaussian phase structure of the laser beams, which shows a phase deviation (Gouy phase shift) $\xi(z) = \arctan z/z_\mathrm{R}$ from a plane wave in direction of the optical axis z, where z_R denotes the Rayleigh range and $z = 0$ the focal point. For a particle traveling along z with velocity v, this phase change results in a frequency shift of $\delta\nu^\mathrm{wf} = v\mathrm{d}\xi(z)/\mathrm{d}z$. From the measured position of the foci and the Rayleigh ranges of both lasers during the beamtime presented here, the shifts are estimated by a Monte-Carlo simulation as $\delta\nu_\mathrm{a}^\mathrm{wf} = (-665 \pm 160)$ kHz for the dye laser and $\delta\nu_\mathrm{p}^\mathrm{wf} = (179 \pm 70)$ kHz for the Ar^+ laser.

Result of the Saturation Spectroscopy

The first two columns of Table 1 comprise all involved frequencies as well as systematic errors from the measurement published in [14]. The difference of the measured Lamb dip position and the value predicted by Special Relativity $\Delta = \nu_\mathrm{a}^\mathrm{exp} - \nu_\mathrm{a}^\mathrm{SR} = -278 \pm 915$ kHz is compatible with zero within the 1 σ uncertainty. The upper limit for deviations from time dilation is thereby further lowered to

$$|\hat{\alpha}| < 2.2 \times 10^{-7} \tag{4}$$

using the β^2 term from (3). The present status of the storage ring experiment thus exceeds previous atomic beam measurements by one order of magnitude. This result is presently limited by the uncertainty of the rest frequency ν_0 [15], which enters (3) quadratically.

4 Outlook

Several further improvements of the control of systematic error sources have been achieved recently. Column 3 of Table 1 shows the expected uncertainties from an experiment employing these new features, which is currently under way. The most significant improvement is achieved by faster chopping as described above. A tribute to the quality of the storage ring spectroscopy is the fact that the limitation at this stage comes from the insufficient knowledge of the rest frame frequency $\Delta\nu_0 = 400$ kHz [15] as it enters in (3) quadratically. Currently, a repetition of the TSR experiment at a low ion velocity ($\beta = 0.03$) is performed

Table 1. Accuracy budget of the saturation spectroscopy; errors are quoted as 1σ, all values in kHz

	Result by Saathoff et al. [14] $\beta = 0.06$		Next generation $\beta = 0.06$	$\beta = 0.03$
	Frequency	1σ error	Error estimates	
Iodine reference line dye	512 671 028 023	152	100	100
Frequency calibration		50	50	100
AOM shift (dye laser) $\delta\nu_{\mathrm{Dye}}^{\mathrm{AOM}}$	414 000	negl.	negl.	negl.
Lamb dip offset to reference $\delta\nu_{\mathrm{L}}$	1 550	460	150	150
Wavefront corr. (dye laser) $\delta\nu_{\mathrm{Dye}}^{\mathrm{wf}}$	−665	160	70	35
Laser-laser angle		40	40	20
Laser-ion angle		10	10	5
Ion beam divergence		10	10	5
Total $\nu_{\mathrm{a}}^{\mathrm{exp}}$	512 671 442 908	517	204	210
$\nu_{\mathrm{p}}^{\mathrm{exp}}$ (incl Laser curvature effect)	582 490 203 442	93	99	122
$^7\mathrm{Li}^+$ rest frequency ν_0 [15]	546 466 918 790	400	100	100
SR prediction $\nu_{\mathrm{a}}^{\mathrm{SR}} = \nu_0^2/\nu_{\mathrm{p}}^{\mathrm{exp}}$	512 671 443 186	755	200	
$\nu_{\mathrm{a}}^{\mathrm{exp}} - \nu_{\mathrm{a}}^{\mathrm{SR}}$	−278	915	300	

in order to replace ν_0. The expected accuracy of this experiment is listed in column 4 of Table 1. Ultimately, we estimate the TSR measurements to provide a limit for $|\hat{\alpha}|$ of 8×10^{-8}.

The investigation of the characteristics of collinear spectroscopy on fast ion beams allows us to scale systematic errors to considerably higher velocities. At the Experimental Storage Ring (ESR) of the Gesellschaft für Schwerionenforschung in Darmstadt, Li^+ can be stored at velocities up to $\beta = 0.45$. Using a beam of, e.g, $\beta = 1/3$, the Doppler-shift leads to wavelengths of $\lambda_{\mathrm{a}} = 776$ nm and $\lambda_{\mathrm{p}} = \lambda_{\mathrm{a}}/2 = 388$ nm, respectively, which can be generated by one laser and its 2^{nd} harmonic. The frequency determination can be accomplished using a self-referenced frequency comb [16]. An estimation based on the experience from the TSR experiment promises an improvement of the sensitivity to $\hat{\alpha}$ into the 10^{-9} range. This experiment is currently being prepared and is anticipated to start in autumn 2005.

Tests of Local Lorentz Invariance are frequently discussed in the framework of the standard model extension by Kostelecký and co-workers [17], which allows for CPT and Lorentz violation that may arise in unifying theories at the Planck scale $E_{\mathrm{p}} \approx 10^{19}$ GeV but may also cause small residual effects in high precision low-energy experiments. Preliminary results by C. Lane [18] indicate a unique sensitivity of Doppler-effect experiments to several parameters in the particle sector of this model. Moreover, their analysis shows that a modified version of our experiment exciting transitions between specific m-sublevels would constrain additional parameters of the model that are not governed by the Mansouri–Sexl test theory. Additionally, the Doppler shift measurements have recently been analyzed in the photon sector [19] of the standard model extension. They

found the Ives-Stilwell experiment to be sensitive to one parameter that was not restricted by other experiments before.

Acknowledgments

We would like to thank H. Buhr, L. A. Carlson, U. Eisenbarth, M. Grieser, K. Horn, H. Krieger, S. Krohn and R. Muñoz Horta for their help in the experiments. Helpful discussions with V. A. Kostelecký, C. Lämmerzahl, C. Lane, and M. Weidemüller are acknowledged.

References

1. H.P. Robertson, Rev. Mod. Phys. **21**, 378 (1949).
2. H.E. Ives and G.R. Stilwell, J. Opt. Soc. Am. **28**, 215 (1938).
3. R. Mansouri and R.U. Sexl, Gen. Rel. Grav. **8**, 497, 515, 809 (1977)
4. A.A. Michelson and E.W. Morley, Am. J. Sci. **34**, 333 (1887).
5. R.J. Kennedy and E.M. Thorndike, Phys. Rev. **42**, 400 (1932).
6. M. Kretzschmar, Z. Phys. A **342**, 463 (1992)
7. R.W. McGowan et al., Phys. Rev. Lett. **70**, 251 (1993).
8. E. Riis et al., Phys. Rev. Lett. **60**, 81 (1988).
9. P. Wolf and G. Petit, Phys. Rev. A **56**, 4405 (1997).
10. D. Habs et al., Nucl. Instr. Meth. B **43**, 390 (1989).
11. J. Kowalski et al., Z. Phys. A **313**,147 (1983).
12. R. Grieser et al., Appl. Phys. B **59**, 127 (1994).
13. G. Saathoff et al., Can. J. Phys. **83**, 425 (2005).
14. G. Saathoff et al., Phys. Rev. Lett. **91**, 190403 (2003).
15. E. Riis et al., Phys. Rev. A **49**, 207 (1994).
16. T. Udem et al., Nature **416**, 233 (2002).
17. D. Colladay and V.A. Kostelecký, Phys. Rev. D **58**, 116002 (1998).
18. C. Lane in: V. A. Kostelecký (ed.), Proceedings of the 3rd Meeting on CPT and Lorentz Symmetry: Bloomington, USA, (World Sci., 2005).
19. P. Wolf et al., Phys. Rev. Lett. **90**, 060402 (2003).

Tests of Lorentz Symmetry in the Spin-Coupling Sector

R.L. Walsworth

Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138, USA
rwalsworth@cfa.harvard.edu

Abstract. An overview is given of recent and ongoing experiments constraining Lorentz violation in the spin-coupling sector, with particular focus on the author's tests of Lorentz symmetry using a ^{129}Xe/^{3}He Zeeman maser and an atomic hydrogen maser.

1 Introduction

Experiments involving spin-polarized systems provide some of the most sensitive tests of Lorentz symmetry. Most commonly, these spin-coupling experiments are modern versions of Hughes-Drever experiments, in which devices related to highly-stable atomic clocks are used to search for a sidereal variation of an atomic Zeeman splitting as the apparatus is rotated and/or boosted by the Earth's motion or by a movable platform. Such experiments usually have several important features:

(1) *High-sensitivity to absolute changes in spin precession frequency*, which generally entails a narrow-bandwidth spin resonance, a large signal-to-noise ratio, and stability over the sidereal modulation period.
(2) *Suppression of sensitivity to magnetic fields*, often by using a co-magnetometer that does not eliminate sensitivity to Lorentz violation.
(3) *Careful engineering to minimize systematics*, e.g., diurnal and seasonal environmental changes for experiments that exploit the Earth's motion to rotate and boost the apparatus.
(4) *A simple spin structure*, to allow a clean interpretation of the experimental results for possible Lorentz-violation of electrons, neutrons, and/or protons.

In the following I provide a brief discussion of six recent and ongoing tests of Lorentz symmetry in the spin-coupling sector, giving particular focus to the experiments with which I am most familiar – i.e., the experiments performed by my group. The results from these experiments are interpreted in terms of the Standard-Model extension (SME) [1], which is reviewed extensively by R. Bluhm in this volume on page 191.

R.L. Walsworth: *Tests of Lorentz Symmetry in the Spin-Coupling Sector*, Lect. Notes Phys. **702**, 493–505 (2006)
DOI 10.1007/3-540-34523-X_18 © Springer-Verlag Berlin Heidelberg 2006

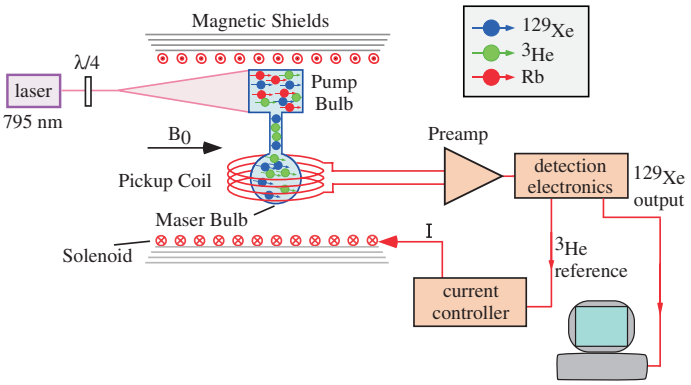


Fig. 1. Schematic of the Harvard-Smithsonian ^{129}Xe/^{3}He Zeeman maser

2 ^{129}Xe/^{3}He maser (Harvard-Smithsonian Center for Astrophysics)

Using a two-species ^{129}Xe/^{3}He Zeeman maser, the author and collaborators at the Harvard-Smithsonian Center for Astrophysics placed a limit on rotation-dependent Lorentz violation involving the neutron of 10^{-31} GeV [2], improving by more than an order of magnitude on the best previous measurement [3, 4]. With the same device we performed the first clean test for the fermion sector of the symmetry of spacetime under boost transformations, placing a limit on boost-dependent Lorentz violation involving the neutron of 10^{-27} GeV [5].

We provide here a brief review of the design and operation of the two-species ^{129}Xe/^{3}He maser. (See the schematic in Fig. 1.) Co-located ensembles of ^{129}Xe and ^{3}He atoms at pressures of hundreds of mbar are held in a double-chamber glass cell placed in a homogeneous magnetic field of ∼1.5 G. Both species have spin-1/2 nuclei and the same sign nuclear magnetic dipole moment, but no higher-order electric or magnetic nuclear multipole moments. In one chamber of the glass cell, the noble gas atoms are nuclear-spin-polarized by spin-exchange collisions with optically-pumped Rb vapor [11]. The noble gas atoms diffuse into the second chamber, which is surrounded by an inductive circuit resonant both at the ^{3}He and ^{129}Xe Zeeman frequencies (4.9 kHz and 1.7 kHz, respectively). For a sufficiently high flux of population-inverted nuclear magnetization, active maser oscillation of both species can be maintained indefinitely.

Due to the generally weak interactions of noble gas atoms with the walls and during atomic collisions, the ^{3}He and ^{129}Xe ensembles can have long Zeeman coherence (T_2) times of hundreds of seconds. It is possible to achieve excellent absolute frequency stability with one of the noble-gas masers by using the second maser as a co-magnetometer. For example, Zeeman frequency measurements with sensitivity of ∼100 nHz are possible with averaging intervals of about an hour. This two-species noble gas maser can also serve as a sensitive NMR gyroscope [12]: the above quoted frequency stability implies a rotation sensitivity of about 0.6 μrad/s averaged over an hour.

In the context of the SME, the neutron – and hence the noble-gas maser – is sensitive to Lorentz violation controlled by the coefficients b_Λ, $d_{\Lambda\Sigma}$, $H_{\Lambda\Sigma}$, and $g_{\Lambda\Sigma\Gamma}$ of the SME [1]. We assume that these coefficients are static and spatially uniform in the Sun frame, at least over the course of a solar year. Thus, the frequencies of the noble-gas masers acquire a time dependence as a consequence of the Earth's rotation and its revolution around the Sun.

In the completed Lorentz-symmetry test, the ^{129}Xe maser was phase-locked to a signal derived from a hydrogen maser in order to stabilize the magnetic field which was oriented along the east-west direction. The leading Lorentz-violating frequency variation of the free-running ^{3}He maser was given by:

$$\delta\nu_{\rm He} = \delta\nu_X \ \sin\ \omega_\oplus T_\oplus + \delta\nu_Y \ \cos\ \omega_\oplus T_\oplus \ , \tag{1}$$

where

$$\begin{aligned} \delta\nu_X &= k\ (\lambda_s + \beta_\oplus(\Lambda_{ss}\sin\Omega_\oplus T + \Lambda_{sc}\cos\Omega_\oplus T)) \ , \\ \delta\nu_Y &= k\ (\lambda_c + \beta_\oplus(\Lambda_{cs}\sin\Omega_\oplus T + \Lambda_{cc}\cos\Omega_\oplus T)) \ . \end{aligned} \tag{2}$$

Here λ_c, λ_s, Λ_{ss}, Λ_{sc}, ... are combinations of Sun-frame SME coefficients mentioned above [5]; $\omega_\oplus$ is the Earth's sidereal angular rotation frequency; $\Omega_\oplus$ is the angular frequency of the Earth's orbital motion; the time $T_\oplus$ is measured in the Sun-centered frame from the beginning of the sidereal day; the time T sets the timescale in the Sun-centered frame (see [5]); $\beta_\oplus$ is the ratio of the Earth's orbital speed to the speed of light; and $k = -8.46 \cdot 10^{32}$ nHz/GeV [2].

We note that (1) and (2) cleanly distinguish the effects of rotation alone (terms proportional to λ_c and λ_s) from the effects of boosts due to the Earth's motion (terms proportional to Λ_{cc}, Λ_{cs}, Λ_{sc}, Λ_{ss}). These equations also indicate that the sensitivity of our experiment to violations of boost-symmetry is reduced by a factor of $\beta_\oplus \simeq 10^{-4}$ with respect to the sensitivity to rotation-symmetry violation.

As discussed in [2] and [5], we acquired noble-gas maser data in four different runs spread over about 13 months (see Fig. 2). Each run lasted about 20 days, and we reversed the direction of the magnetic field about halfway through each run to help distinguish possible Lorentz-violating effects from diurnal systematic variations. We fit this data to (1). Figure 2 shows, for each run, the mean values we determined for $\delta\nu_X$ and $\delta\nu_Y$, the amplitudes of sidereal-day modulations of

Table 1. Bounds from the competed noble gas maser experiment on 17 SME coefficients among the 44 coefficients describing possible leading-order Lorentz- and CPT-violating coupling of the neutron

SME Coefficients	GeV
$\tilde{b}_X, \tilde{b}_Y$	$[10^{-31}]$
$\tilde{d}_X, \tilde{d}_Y, \tilde{g}_{DX}, \tilde{g}_{DY}$	$[10^{-28}]$
$\tilde{b}_T, \tilde{d}_{XY}, \tilde{d}_{YZ}, \tilde{d}_+, \tilde{d}_-, \tilde{d}_Q, \tilde{g}_T, \tilde{g}_c, \tilde{H}_{XT}, \tilde{H}_{YT}, \tilde{H}_{ZT}$	$[10^{-27}]$

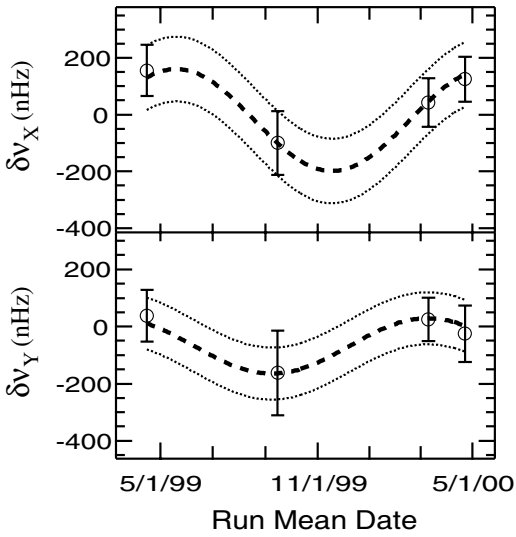


Fig. 2. Time course of the mean values of $\delta\nu_X$ and $\delta\nu_Y$ from the completed noble gas maser Lorentz symmetry test. For each plot the dashed line is the best fit obtained from (2), using the fit parameters λ_c, λ_s, Λ_{cc}, Λ_{cs}, Λ_{sc}, Λ_{ss}. Dotted lines indicate the $1\,\sigma$ confidence bands for the fit model

the ^{3}He-maser frequency due to Lorentz-violating coefficients in the $\hat{X}$ and $\hat{Y}$ directions (Sun-centered frame). For each run, $\delta\nu_X$ and $\delta\nu_Y$ correspond to a very good approximation to a single high-precision measurement of the X and Y components of $\delta\nu_{He}$ performed at the run's mean time.

Next, we fit the experimental values of $\delta\nu_X$, $\delta\nu_Y$ to (2), thus obtaining the fit shown graphically in Fig. 2, and the corresponding bounds on the SME coefficients of Table 1. We treated all fit parameters as independent and we extracted energy bounds for SME coefficients disregarding the possibility of accidental mutual cancellations. This analysis yielded no significant violation of rotation invariance with a limit of about 70 nHz on the magnitude of the daily sidereal variation in the ^{3}He-maser frequency and no significant violation of boost invariance, with a limit of about 150 nHz on the magnitude of an annual modulation of the daily sidereal variation. This experiment was not limited by systematic effects.

We expect an order of magnitude or more improved sensitivity to Lorentz violation of the neutron using a reengineered version of our ^{129}Xe/^{3}He maser – a project currently underway. The new device has been designed to improve the frequency stability of the noble gas masers, which limits the current sensitivity to Lorentz violation. Improved temperature control of the pump and maser regions, better co-magnetometry through optimized gas pressures and cell geometry, and the use of a narrow spectrum laser for optical pumping should help achieve this goal. Further improvements in sensitivity may be possible with a ^{21}Ne/^{3}He

Zeeman maser [13], with masers located on a rotating table, or with space-based clocks [14].

3 Hydrogen Maser (Harvard-Smithsonian Center for Astrophysics)

The author and collaborators employed atomic hydrogen masers to set an improved clean limit on rotation-violation of the proton, at the level of nearly 10^{-27} GeV [6].

Hydrogen masers operate on the $\Delta F = 1$, $\Delta m_F = 0$ hyperfine transition in the ground state of atomic hydrogen [15]. Hydrogen molecules are dissociated into atoms in an RF discharge, and the atoms are state selected via a hexapole magnet (Fig. 3). The high field seeking states, ($F = 1$, $m_F = +1, 0$) are focused into a Teflon coated cell which resides in a microwave cavity resonant with the $\Delta F = 1$ transition at 1420 MHz. The $F = 1$, $m_F = 0$ atoms are stimulated to make a transition to the $F = 0$ state by the field of the cavity. A static magnetic field of ∼1 milligauss is applied to maintain the quantization axis of the H atoms.

The hydrogen transitions most sensitive to potential Lorentz violations are the $F = 1$, $\Delta m_F = \pm 1$ Zeeman transitions, which are effectively degenerate (with frequency $\nu_Z < 1$ kHz) for the typical static magnetic field. We utilize a double resonance technique to measure the Zeeman frequency with a precision of ∼1 mHz [6, 16, 17]. We apply a weak oscillating magnetic field perpendicular to the static field, and slowly sweep the oscillating field's frequency through the Zeeman transition. This audio-frequency driving field couples the three sublevels

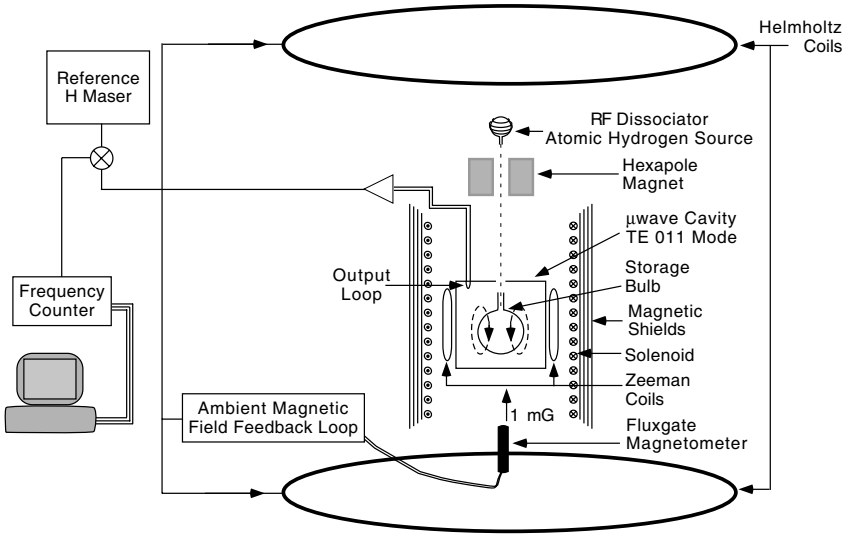


Fig. 3. Schematic of the Harvard-Smithsonian H maser in the ambient field stabilization loop used for Lorentz symmetry tests

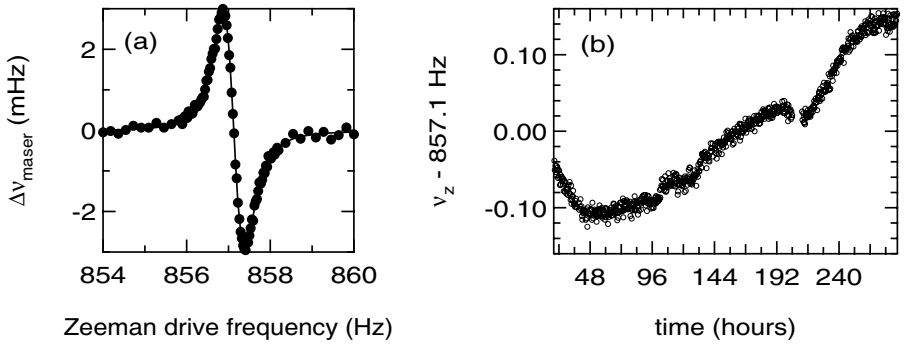


Fig. 4. (**a**) An example of a double resonance measurement of the $F = 1$, $\Delta m_F = \pm 1$ Zeeman frequency in the hydrogen maser. The change from the unperturbed maser frequency is plotted versus the driving field frequency. (**b**) Zeeman frequency data from 11 days of the completed Lorentz symmetry test using the H maser

of the $F = 1$ manifold of the H atoms. Provided a population difference exists between the $m_F = \pm 1$ states, the energy of the $m_F = 0$ state is altered by this coupling, thus shifting the measured maser frequency in a carefully analyzed manner [6, 16, 17] described by a dispersive shape (Fig. 4a). Importantly, the maser frequency is unchanged when the driving field is exactly equal to the Zeeman frequency. Therefore, we determine the Zeeman frequency by measuring the driving field frequency at which the maser frequency in the presence of the driving field is equal to the unperturbed maser frequency.

We employ an active stabilization system to cancel external magnetic field fluctuations (Fig. 3). A fluxgate magnetometer placed within the maser's outer magnetic shield controls large (2.4 m dia.) Helmholtz coils surrounding the maser via a feedback loop to maintain a constant ambient field. This feedback loop reduces the fluctuations at the sidereal frequency to below the equivalent of 1 μHz on the Zeeman frequency at the location of the magnetometer.

In the completed H maser Lorentz symmetry test, the hydrogen Zeeman frequency was measured for 32 days using the double resonance technique. During data taking, the maser remained in a closed, temperature controlled room to reduce potential systematics from thermal drifts which might be expected to have 24 hour periodicities. The feedback system also maintained a constant ambient magnetic field. Each Zeeman measurement took approximately 20 minutes to acquire and was subsequently fit to extract a Zeeman frequency (Fig. 4a). Also monitored were maser amplitude, residual magnetic field fluctuation, ambient temperature, and current through the solenoidal coil which determines the Zeeman frequency (Fig. 3).

The data were then fit to extract the sidereal-period sinusoidal variation of the Zeeman frequency. (See Fig. 4b for an example of 11 days of data.) In addition to the sinusoid, piecewise linear terms (whose slopes were allowed to vary independently for each day) were used to model the slow remnant drift of the

Zeeman frequency. No significant sidereal-day-period variation of the hydrogen $F = 1$, $\Delta m_F = \pm 1$ Zeeman frequency was observed, setting a bound on the magnitude of such a variation of $\delta\nu_Z^H \leq 0.37$ mHz (one-sigma level). This experiment was not limited by systematic effects.

In the context of the SME, the H maser measurement constrains Lorentz (rotation) violation of the proton parameter $|\tilde{b}_{x,y}^p| \leq 2 \cdot 10^{-27}$ GeV at the one sigma level (Earth-centered frame), given the much more stringent limits on Lorentz violation of the electron set with spin-torsion pendula (see discussion below).

We expect that the sensitivity of the H maser Lorentz symmetry test can be improved by more than an order of magnitude through technical upgrades to the maser's thermal and magnetic field systems; better environmental control of the room housing the maser; and a longer period of data acquisition. Such improvements are underway.

4 Spin-Torsion Pendula (University of Washington and Tsing-Hua University)

In separate research efforts, groups at the University of Washington (Adelberger, Gundlach, Heckel et al.) and Tsing-Hua University in Taiwan (Hou, Ni, and Li) each employ a spin-torsion pendulum to search for Lorentz violation of the electron. The Univ. Washington effort is the most established, and has achieved sensitivity to certain SME rotation-violation parameters for the electron ($\tilde{b}_{x,y}^e$) at the level of 10^{-30} GeV [18]. The Taiwan experiment, which is not reviewed here, has set somewhat less-stringent bounds for rotation-violation of the electron: $|\tilde{b}_{x,y}^e| < 3 \times 10^{-29}$ GeV [19]. (Here, an Earth-centered frame is used, with the $\hat{z}$ axis taken to lie along the rotational north pole.)

The Univ. Washington group employs a pendulum test mass with a large net electron spin dipole moment but a small magnetic moment, thereby enabling a sensitive search for Lorentz-violating spin-coupling with minimal confounding magnetic interactions. The pendulum has a toroidal (ring) geometry constructed fromeight sections of two different kinds of permanent magnets: four sections of an aluminum-nickel-cobalt-iron alloy (Alnico) and four sections of a samarium-cobalt magnet ($SmCo_5$). The magnetization in Alnico comes primarily from electron spin while the magnetization in $SmCo_5$ is produced both by electron spin and the orbital angular momentum of the Sm ions. The octagon magnet ring is assembled with the Alnico pieces on one side and the $SmCo_5$ pieces on the other. The Alnico is then magnetized to the same degree as the $SmCo_5$, with the result that the magnetization runs azimuthally within the ring. Thus there is both near perfect cancellation of the net magnetic moment of the pendulum and also a net electron spin excess within the Alnico side. Four such rings are stacked in an ABBA pattern with their net electron spin axes aligned, to form a pendulum with a net spin dipole of $(8 \pm 1) \times 10^{22}$ electron spins pointing perpendicular to the central axis of the magnet stack. The basic layout and principle is illustrated in Fig. 5.

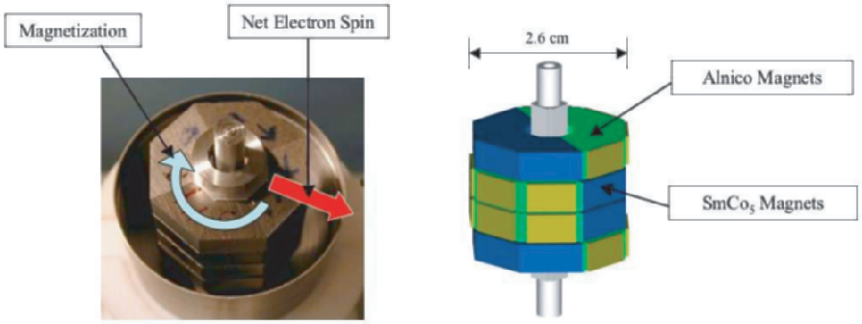


Fig. 5. (*Left*) Top of the stack of four magnet rings that consitute the University of Washington spin pendulum. Superimposed arrows indicate the directions of the (azimuthal) magnetization and the large net electron spin. (*Right*) Stack of 4 magnet rings in an ABBA pattern, which constitute the spin pendulum. Small plates are added to the Alnico sections to give them the same mass as the SmCo sections

The spin pendulum is suspended from a tungsten torsion fiber and centered within four layers of high permeability magnetic shields. The pendulum and shields are housed within a vacuum chamber, which is mounted on a turntable that rotates at a constant rate of approximately 4 rev/hr. A feedback loop locks the output of a precise rotary encoder attached to the rotating vacuum chamber to the frequency of a crystal oscillator to ensure a constant rotation rate. Diode laser light is doubly reflected from one of four mirrors mounted on the pendulum, and the reflected beam is focused onto a linear position sensitive photodiode to monitor the angular position of the pendulum, with a sensitivity of a few nanoradians. The apparatus is surrounded by three-axis Helmholtz coils and carefully positioned masses to reduce magnetic and gravitational field gradients. As the vacuum chamber and pendulum within it rotate relative to the laboratory, any Lorentz-violating field coupling to electron spin induces a torque on the pendulum modulated at the rotation period of the turntable. In about one hundred days of preliminary data acquisition in 2004/05, no such rotation-violation was observed, at the level of approximately 10^{-30} GeV; also, no evidence has been found of systematic error, e.g., from temperature effects, magnetic coupling to the pendulum, and gravitational gradients [18]. The Univ. Washington group is currently acquiring an additional year of data, with an expected sensitivity of $|\tilde{b}^e_{x,y}| \approx 3 \times 10^{-31}$ GeV, as well as potential sensitivity of $\sim 10^{-27}$ GeV to boost symmetry violation in the electron sector.

^{199}Hg/^{133}Cs co-magnetometer (Amherst College)

An experiment is underway at Amherst College (Hunter and collaborators) to compare the precession frequencies of ^{199}Hg and ^{133}Cs magnetometers as a function of the sidereal rotation and boost of the system's quantization axis (set by a weak applied magnetic field). This experiment is a successor to a high-precision test of Lorentz symmetry performed in 1995 [3], which provided the most sensitive limits for rotation-symmetry-violation of the neutron ($\sim 10^{-30}$ GeV) prior to the noble gas maser experiment discussed above. Both the 1995 and current Amherst experiments employ optically-pumped Hg and Cs magnetometer cells, which are specially prepared to provide long spin-relaxation times. Each cell is probed independently as a light-absorption oscillator, which allows sensitive measurement of the associated spin precession frequency and thus the average magnetic field at each cell. A series of coils and several layers of surrounding magnetic shields create a very homogeneous magnetic field ($\sim$5 mG) across all magnetometer cells. Typically, one Cs cell is placed between two Hg cells, such that the average magnetic field at the Hg cells is the same as the field at the Cs cell if there is any remnant, linear gradient of the magnetic field. See Fig. 6. The signature of Lorentz-violation is a modulation of the Hg and/or Cs spin precession frequencies as the quantization axis rotates or boosts sidereally. ^{199}Hg has no electron spin and nuclear spin 1/2 (with the dominant contribution coming from a valence neutron), whereas ^{133}Cs has electron spin 1/2 and nuclear spin 7/2 (with complicated nuclear structure and contributions from both neutrons and protons).

The 1995 experiment had excellent short-term sensitivity to effects coupling to the nuclear and electronic spins of ^{199}Hg and ^{133}Cs; however, the experiment relied on the Earth's rotation and the long-term stability was compromised by its environmental sensitivity and drift in operational parameters. Improvements

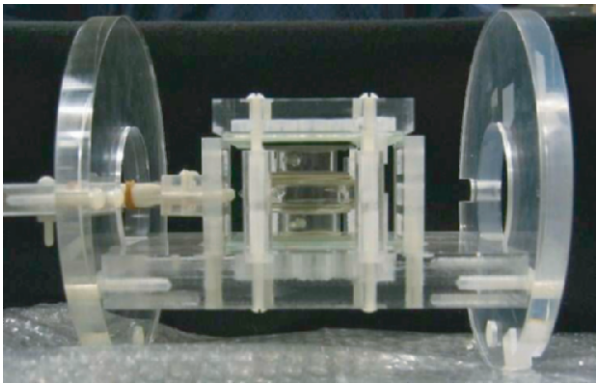


Fig. 6. Magnetometer cell assembly for the Lorentz symmetry test currently being performed at Amherst College. One Cs cell resides between two Hg cells in a vertical stack at the center of the assembly

in the new experiment include a solid state laser system, which has replaced the earlier experiment's Hg discharge lamp. The intensity, frequency and polarization of the laser light are carefully regulated and should result in improved long-term stability of the Hg magnetometer. In addition, the entire new apparatus (save the first two stages of the Hg laser) has been mounted on a rotation table, and assembled vertically in the lab such that the magnetic shields rotate about their axis of symmetry. The apparatus is rotated between two positions approximately every 6 minutes to take advantage of the Hg/Cs co-magnetometer's excellent short-term sensitivity. These positions correspond to the applied magnetic field pointing either perpendicular or parallel to the rotation axis of the Earth. The parallel position provides a fixed reference direction (i.e., with no leading-order Lorentz-violation effect). Taking the difference between measurements made in these two positions allows removal of contributions from slow drifts in the apparatus over a sidereal-day-period, which is the signature of a preferred spatial direction. These technical upgrades are expected to enable two orders of magnitude greater sensitivity to Lorentz violation than the 1995 measurement, with the greatest sensitivity being for the neutron.

5 K/^{3}He Co-Magnetometer (Princeton University)

At Princeton University, Romalis and collaborators are currently pursuing a promising Lorentz symmetry test using co-located K and ^{3}He atoms, which together act as a zero-field, self-compensating magnetometer and provide excellent sensitivity to non-magnetic fields that couple differently to the K and ^{3}He spins [20–22]. Optical pumping by a strong pump laser spin-polarizes the K atoms along the longitudinal axis (the direction of the pump beam's propagation). Any small transverse component of the K spin-polarization (e.g., induced by Lorentz-violating fields) is measured by optical rotation of a weak probe laser beam directed orthogonal to the pump beam. The K atoms are maintained at relatively high-density ($\sim 10^{13}\,\mathrm{cm}^{-3}$) and near-zero magnetic field in order (i) to eliminate decoherence of K spin-precession due to K-K spin-exchange collisions [23], and (ii) to provide excellent signal-to-noise for measurement of transverse K spin-polarization. K-^{3}He collisions polarize the ^{3}He nuclei through a transient hyperfine interaction, and also enhance the dipolar interaction between K and ^{3}He atoms. The spin-polarized ^{3}He gas is at high density ($\sim 10^{20}$ cm^{-3}), and thus develops a significant magnetization which imposes an effective magnetic field $\sim$1 mG on the K atoms. A longitudinal magnetic field is applied to cancel this effective ^{3}He field, such that the total magnetic field experienced by the K atoms vanishes. The ^{3}He magnetization adiabatically follows slow changes in the transverse magnetic field, thereby compensating for magnetic field drifts and maintaining the K atoms in a highly-stable zero field. Thus, transversely-directed Lorentz-violating fields that couple differently to the K and ^{3}He spins will induce a small transverse K spin-polarization, both through a direct torque on the K spins, and through an induced misalignment of the ^{3}He magnetization

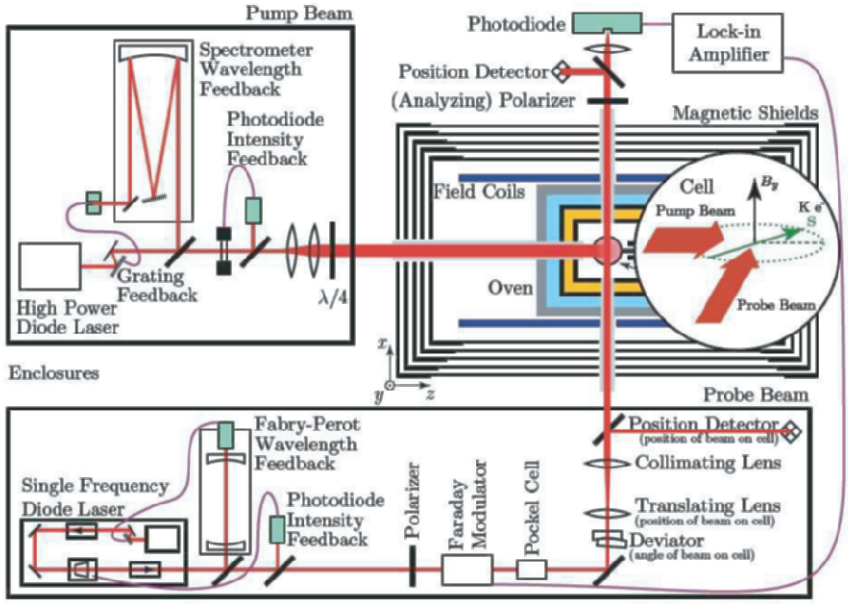


Fig. 7. Schematic of the K/^{3}He co-magnetometer at Princeton University currently being used in a high-precision test of Lorentz symmetry

and the applied magnetic field with a resultant torque of the net magnetic field on the K spins. The Princeton group relies on the Earth's motion to rotate and boost the apparatus; hence signals for Lorentz violation would include a sidereal day and year period modulation of the transverse K spin-polarization.

A schematic of the Princeton experiment is shown in Fig. 7. A near-spherical 2.5 cm diameter glass cell holds the K and ^{3}He atoms, along with 50 Torr of N_2 gas to inhibit radiation trapping in the optical pumping process. A blown-air oven heats the magnetometer cell to about 175°C to create the appropriate K density. Five-layer magnetic shields provide isolation of $\sim 10^6$ to external magnetic fields, and precision coils create a uniform longitudinal magnetic field (to balance the effective ^{3}He magnetic field experienced by the K atoms). A 770 nm, 1 W broad area diode laser provides optical pumping on the K $D1$ line; a single-mode diode laser with a tapered amplifier produces the 50 mW probe beam, which is linearly polarized and detuned about 1 nm to the blue of the $D1$ line. Both the pump and probe beams have active control of their wavelength and intensity. The probe beam polarization is weakly modulated before passing through the magnetometer cell, and lock-in detection is used to measure any optical rotation induced by a transverse component of the K spin-polarization. With this system, the Princeton group has achieved short-term sensitivity of $\sim 10^{-31}$ GeV for anomalous fields coupling to the neutron spin, and $\sim 10^{-28}$ GeV for anomalous fields coupling to the electron spin. To realize precision tests of rotation and boost symmetry, this excellent short-term sensitivity must be re-

alized over periods of a sidereal day and year. The Princeton group is currently working to stabilize long-term drifts in system parameters and to characterize systematic effects. For example, small motions of the pump and probe beams, driven by drifts in environmental temperature, have proven to be a primary source of systematic error. Once these effects are reduced or controlled, it is expected that the K/^{3}He co-magnetometer will enable sensitivity to rotation (boost) violation up to $\sim 10^{-33}$ GeV ($\sim 10^{-29}$ GeV) for the neutron, and $\sim 10^{-30}$ GeV ($\sim 10^{-26}$ GeV) for the electron.

References

1. V.A. Kostelecký, C.D. Lane: Phys. Rev. D **60**, 116010 (1999).
2. D. Bear, R.E. Stoner, R.L. Walsworth, V.A. Kostelecký, and C.D. Lane : Phys. Rev. Lett. **85**, 5038 (2000); *ibid.*, **89**, 209902 (2002).
3. C.J. Berglund, L.R. Hunter, D. Krause, Jr., E.O. Prigge, M.S. Ronfeldt, and S.K. Lamoreaux: Phys. Rev. Lett. **75**, 1879 (1995).
4. L.R. Hunter, C.J. Berglund, M.S. Ronfeldt et al: A Test of Local Lorentz Invariance Using Hg and Cs Magnetometers. In: *CPT and Lorentz Symmetry*, ed by V.A. Kostelecký (World Scientific, Singapore, 1999) pp 180–186.
5. F. Canè, D. Bear, D.F. Phillips, M.S. Rosen, C.L. Smallwood, R.E. Stoner, R.L. Walsworth, and V.A. Kostelecký: Phys. Rev. Lett. **93**, 230801 (2004).
6. D.F. Phillips, M.A. Humphrey, E.M. Mattison, R.E. Stoner, R.F.C. Vessot, and R.L. Walsworth: Phys. Rev. D **63**, 111101 (2001). M.A. Humphrey, D.F. Phillips, E.M. Mattison, R.F.C. Vessot, R.E. Stoner, and R.L. Walsworth: Phys. Rev. A **68**, 063807 (2003).
7. D. Colladay, V.A. Kostelecký: Phys. Rev. D **55**, 6760 (1997); **58**, 116002 (1998). V.A. Kostelecký, R. Lehnert: Phys. Rev. D **63**, 065008 (2001). V.A. Kostelecký: Phys. Rev. D **69**, 105009 (2004).
8. O.W. Greenberg: Phys. Rev. Lett. **89**, 231602 (2002); Phys. Lett. B **567**, 179 (2003).
9. V.A. Kostelecký, S. Samuel: Phys. Rev. D **39**, 683 (1989); Phys. Rev. Lett. **63**, 224 (1989); Phys. Rev. D **40**, 1886 (1989). V.A. Kostelecký, R. Potting: Nucl. Phys. B **359**, 545 (1991); Phys. Rev. D **51**, 3923 (1995).
10. For reviews of approaches to Lorentz and CPT violation, see, for example, *CPT and Lorentz Symmetry I, II, III*, ed by V.A. Kostelecký (World Scientific, Singapore, 1999, 2002, 2004).
11. T.E. Chupp, E.R. Oteiza, J.M. Richardson, and T.R. White: Phys. Rev. A **38**, 3998 (1988). G.D. Cates, R.J. Fitzgerald, A.S. Barton, P. Bogorad, M. Gatzke, N.R. Newbury, and B. Saam: Phys. Rev. A **45**, 4631 (1992).
12. K.F. Woodman et al: J. Navig. **40**, 366 (1987).
13. R.E. Stoner, R.L. Walsworth: Phys. Rev. A **66**, 032704 (2002).
14. R. Bluhm, V.A. Kostelecký, C.D. Lane, and N. Russell: Phys. Rev. Lett. **88**, 090801 (2002).
15. D. Kleppner, H.M. Goldenberg, N.F. Ramsey: Phys. Rev. **126**, 603 (1962). D. Kleppner, H.C. Berg, S.B. Crampton, N.F. Ramsey, R.F.C. Vessot, H.E. Peters, and J. Vanier: Phys. Rev. **138**, A972 (1965).
16. H.G. Andresen: Z. Physik **210**, 113 (1968).
17. M.A. Humphrey, D.F. Phillips, R.L. Walsworth: Phys. Rev. A **62**, 063405 (2000).

18. B.R. Heckel: Torsion Balance Test of Lorentz Symmetry Violation. In: *CPT and Lorentz Symmetry III*, ed by V.A. Kostelecký (World Scientific, Singapore, 2004) pp 133–140.
19. L.-S. Hou, W.-T. Ni, Y.-C.M. Li: Phys. Rev. Lett. **90**, 201101 (2003).
20. T.W. Kornack, M.V. Romalis: Operation of the K-^{3}He Self-Compensating Co-Magnetometer for a Test of Lorentz Symmetry. In: *CPT and Lorentz Symmetry III*, ed by V. A. Kostelecký (World Scientific, Singapore, 2004) pp 57–70.
21. T.W. Kornack, M.V. Romalis: Phys. Rev. Lett. **89**, 253002 (2002).
22. J.C. Allred, R.N. Lyman, T.W. Kornack et al: Phys. Rev. Lett. **89**, 130801 (2002).
23. Because spin-exchange collisions preserve the colliding atoms' total angular momentum, the spin decoherence effects of such collisions are eliminated in the limit that the collision frequency is much greater than the Larmor spin-precession frequency (i.e., the Zeeman frequency) in an applied magnetic field. See: W. Happer, H. Tang: Phys. Rev. Lett. **31**, 273 (1973) and W. Happer, A.C. Tam: Phys. Rev. A **16**, 1877 (1977).

Do Evanescent Modes Violate Relativistic Causality?

G. Nimtz

II. Physikalisches Institut, Universität zu Köln, Zülpicher Str. 77, 50937 Köln, Germany
G.Nimtz@uni-koeln.de

Abstract. Time dependent experiments with evanescent modes (photonic tunneling) can be performed with high precision and at a macroscopic scale with microwaves in the range of meters or in the infrared regime in the range of centimeters. The infrared technology is the present day digital signal processing and transmission. Superluminal (faster than light) signal transmission by evanescent modes was shown by Enders and Nimtz already 1992 [1].

Evanescent modes are solutions of the Helmholtz equation with imaginary wave vectors which are equivalent to the tunneling solutions of the Schrödinger equation. Experiments of transmission and of partial reflection of microwaves by photonic potential barriers revealed superluminal signal velocity of evanescent modes. The effect is a near field phenomenon and violates the relativistic causality.

In this contribution superluminal experiments are introduced and explained.

1 Introduction

During the last decade much research and arguing was devoted to superluminal signal velocity $v_{\text{signal}} > c$, where c is the vacuum velocity of light [2–5]. I am not talking about the phase velocity $\mathbf{v}_{\text{phase}}$, which exceeds in several media the velocity c of light. The relevant *signal velocity* is in charge of the transmission of a *defined cause and subsequent effect*. Actually, the near field phenomena evanescent modes and tunneling represent the exception of $v_{\text{signal}} \leq c$.

In this lecture I present experiments on superluminal signal velocity of evanescent modes. Evanescent modes are solutions of the Helmholtz equation. Like wave mechanical tunneling functions these special solutions are characterized by a purely imaginary wave number. The wave number represents 2π times the reciprocal wavelength. Accordingly evanescent modes do not have a real wave length and the phase time approach conjectured the observed instantaneous field spreading of evanescent modes. However, a superluminal signal velocity does not violate the *primitive causality* (the effect cannot precede the cause). But a superluminal signal velocity violates the *relativistic causality*, often called Einstein causality: no signal can propagate with a velocity greater than c. *Experiments*

G. Nimtz: *Do Evanescent Modes violate Relativistic Causality?*, Lect. Notes Phys. **702**, 506–531 (2006)
DOI 10.1007/3-540-34523-X_19 © Springer-Verlag Berlin Heidelberg 2006

show that detectors receive signals transmitted by evanescent modes earlier than signals, which traveled the same distance in vacuum. For instance the detector makes earlier click in the case of a tunneled digital signal or receives a tunneled melody before detecting the air born one. Even though as explained in Sect. 8 the design of a time machine is still not possible by signaling with superluminal evanescent modes.

Studies with evanescent waves were stimulated in order to obtain analogous experimental data on quantum mechanical tunneling time. Tunneling represents the quantum mechanical analogy to the electromagnetic evanescent modes [6]. As there have been no experimental data on quantum mechanical tunneling time available, the propagation time of evanescent modes was studied, which is easier to measure than particle tunneling time. More over in the case of electron tunneling in semiconductor devices there are present time consuming parasitical Coulomb interactions which determine the measured tunneling time.

The tunneling time is of the order of the reciprocal frequency of the wave packet [13, 15]. This time is spent at the entrance boundary as will be shown in Sect. 6. From an experimental point of view the transit time for a wave packet propagating through a barrier is measured as the interval between the arrivals of the signal envelope at the two ends of that region.

An example of evanescent digital signals transmitted with microwaves is displayed in Fig. 1. The half width (the time duration at half the maximum intensity) represents the number of digits. To make a comparison the small tunneled wave packet is amplified by about a factor of 10 000 (i.e. 40 dB), however, re-

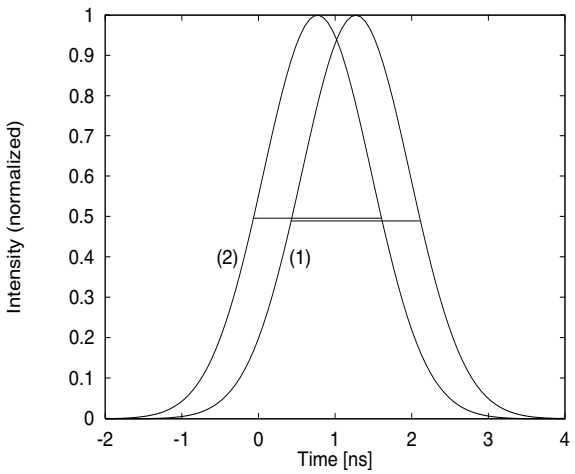


Fig. 1. Intensity vs time of a microwave pulse (2), which has tunneled at superluminal velocity through a photonic barrier in a wave-guide of 114.2 mm length. For comparison the tunneled, digital signal is normalized with a pulse (1), which propagated through a normal waveguide of the same length. The tunneled digital signal traveled at a speed of $4.7\,c$ [7]. The halfwidth (*solid line*) of the pulse equals the number of digits, i.e. it represents the signal

member that a signal is independent of intensity as its intensity exceeds that of the thermal noise as discussed in Sect. 7. The evanescent pulse displayed in Fig. 1 traveled with a superluminal velocity of $4.7\,c$.

Incidentally, $v_{\rm signal} > c$ does not occur in experiments based on a near to resonance interaction with a Lorentz–Lorenz like oscillator. This oscillator is the paradigm of particle polarization in electric fields [8]. In those experiments pulses can display even a negative group velocity [9]. But only the peak of the pulse traveled at a negative group velocity and not the whole envelope of the signal. The interacting field distribution of the signal was reshaped and the signal envelope traveled at subluminal velocity.

At present there is much ado about quantum teleportation [10]. As those sophisticated quantum mechanical experiments include a classical communication channel the signal velocity becomes $v_{\rm signal} \leq c$ finally. Incidentally, teleportation is a technique applied in telecommunication for a long time, where sound (phonons) are transformed into electromagnetic waves (photons), which travel a million times faster than sound and the receiver transformed the electromagnetic waves back into the slow but understandable sound. For the time being evanescent modes and tunneling seem to represent the only mechanisms to achieve superluminal signal velocities.

In the following some elementary quantities and relations are reminded. The propagation of waves $\psi \sim e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)}$ in space is described by the a relation connecting the wave number $\mathbf{k}$ or, equivalently, the wavelength with the angular frequency ω

$$k = k(\omega) = k_0 \cdot n(\omega)\,, \qquad \lambda(n) = \lambda_0/n(\omega)\,. \tag{1}$$

Here k_0 is the wave number and λ_0 the wavelength of waves in vacuum which are related $k_0 = 2\pi/\lambda_0$. Furthermore, $n(\omega) = n'(\omega) - in''(\omega)$ is the refractive index n we are familiar with from Snellius' law. The quantities n' and n'' real and imaginary parts of the refractive index of the medium in question. Both quantities, k and n are in general complex functions of frequency. The imaginary parts describe the attenuation or amplification of waves. The attenuation may be caused either by dissipation or by reflection. Waves with purely imaginary refractive index $n(\omega)$ and wave number $k(n)$ are called *evanescent modes*.

2 Wave Propagation

2.1 Maxwell and Schrödinger Equations

For electromagnetic waves and hence for photons, the propagation of waves can be described by the Maxwell equations and for massive particles by the Klein–Gorden, the Dirac or, in the non-relativistic regime, by the Schrödinger equation.

The Maxwell equations in media characterized by some refractive index $n = \sqrt{\mu\epsilon}$, where μ and ϵ are the relative permeability and the relative permittivity, lead to the wave equation

$$-\nabla^2\phi(\mathbf{x},t) + \frac{n^2}{c^2}\frac{\partial^2}{\partial t^2}\phi(\mathbf{x},t) = 0\,, \tag{2}$$

ϕ being any component of the electrical and the magnetic fields. In vacuum characterized by $n = 1$ waves propagates with the velocity $c = (\mu_0\epsilon_0)^{-1/2}$, where μ_0 and ϵ_0 are the permeability and the permittivity, respectively.

If we describe phenomena periodic in time with frequency $\nu = \omega/(2\pi)$,

$$\phi(\mathbf{x}, t) = \phi_x(\mathbf{x})e^{i\omega t}\,, \tag{3}$$

then the wave equation reduces to the Helmholtz equation

$$\nabla^2\phi_x(\mathbf{x}) + \frac{n^2\omega^2}{c^2}\phi_x(\mathbf{x}) = 0\,. \tag{4}$$

As usual, this equation will be solved by a plane wave ansatz

$$\phi_x(\mathbf{x}) = \phi_0 e^{-i\mathbf{k}\cdot\mathbf{x}}\,, \tag{5}$$

what leads to a relation between the wave number and the refractive index

$$k^2 = \frac{n^2\omega^2}{c^2} = k_0^2 n^2 = k_0^2\epsilon\mu\,, \tag{6}$$

where k_0 is the wave number in free space. If k and, thus, n are purely imaginary then the solution is called an *evanescent mode*. The imaginary wave number is usually expressed by κ. In Sect. 3 we will discuss three popular examples, where these special solutions occur.

Similar features can be found for the stationary Schrödinger equation

$$E\psi(\mathbf{x}) = -\frac{\hbar^2}{2m}\nabla^2\psi(\mathbf{x}) + U(\mathbf{x})\psi(\mathbf{x})\,, \tag{7}$$

where E is the energy of the stationary state, m is the mass of the particle and $U(\mathbf{x})$ is a position-dependent potential, the barrier potential, for example. This relation is mathematically equivalent to the Helmholtz equation

$$\nabla^2\psi(\mathbf{x}) + \frac{2m}{\hbar^2}(E - U(\mathbf{x}))\psi(\mathbf{x}) = 0\,. \tag{8}$$

Again, a plane wave ansatz

$$\psi(\mathbf{x}) = \psi_0 e^{-i\mathbf{k}\cdot\mathbf{x}} \tag{9}$$

yields for the wave number k

$$k^2 = \frac{2m}{\hbar^2}(E - U) = k_0^2 - \frac{2mU}{\hbar^2}\,, \tag{10}$$

where $k_0^2 = 2mE/\hbar^2$ is the wave vector at infinity, where U is assumed to vanish. Particles in regions for which $E < U$, that is, inside the potential barrier, are quantum analogues of evanescent modes. Obviously, for the electromagnetic evanescent modes the refractive index plays the role of the potential in the wave mechanical tunneling.

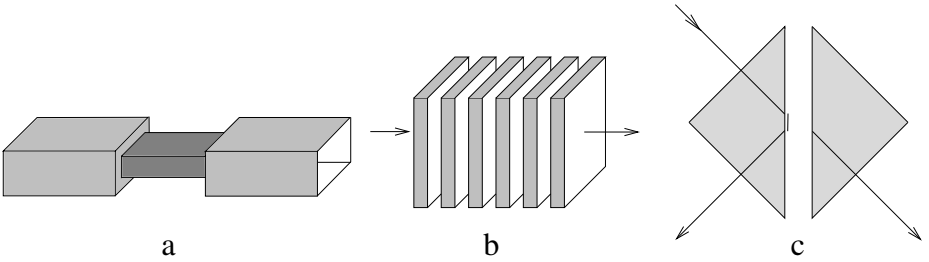


Fig. 2. Sketch of three photonic barriers. (**a**) illustrates an undersized wave guide (the central part of the wave guide has a cross section being smaller than half the wavelength in both directions perpendicular to propagation), (**b**) a 1-dimensional photonic lattice (periodic dielectric hetero structure), and (**c**) the frustrated total internal reflection of a double prism, where total reflection takes place at the boundary from a denser (the first prism with refractive index n_1) to a lesser dielectric medium (with refractive index n_2)

3 Photonic Barriers, Examples of Evanescent Modes

Prominent examples of evanescent modes are found in undersized wave guides (both dimensions of the guide cross section are smaller than half the vacuum wavelength), in the forbidden frequency bands of periodic dielectric hetero-structures (photonic lattice), and with double prisms in the case of frustrated total internal reflection [3,5]. The three examples are illustrated in Fig. 2. Dielectric lattices are analogous to electronic lattices of semiconductors with forbidden energy gaps. As seen below the square number of the imaginary refractive index n''^2 corresponds to a negative effective potential $E-U$ in the Schrödinger equation. Each of the three barriers introduced in Fig. 2 have a different dispersion relation of the wave number $k(\omega)$, of the refractive index $n(\omega)$, and then of the transmissivity $T(\omega)$.

3.1 Undersized Waveguide

Figure 2a displays an undersized waveguide with the long side a of the waveguide cross section which is mounted between two properly sized parts. From the wave equation (2) one can easily determine the wave vector

$$k = k_0\sqrt{1-\left(\frac{\lambda_0}{\lambda_{\mathrm{cutoff}}}\right)^2} = k_0\sqrt{1-\left(\frac{\omega_{\mathrm{cutoff}}}{\omega_0}\right)^2} = k_0 n(\omega)\ , \qquad (11)$$

where we introduced the cutoff wavelength $\lambda_{\mathrm{cutoff}} = 2a$ which is related to the angular cutoff frequency $\omega_{\mathrm{cutoff}} = \pi c/a$. Below the cutoff frequency or above the cutoff wavelength the waveguide wave propagation is prohibited since in that case the wave number $n(\omega)$ becomes imaginary. Then the solution represents an evanescent mode. The intensity of this evanescent mode decreases by $1/e$ at a distance a/π: The field does not propagate and dies off rapidly with distance.

3.2 Photonic Lattice

The photonic lattice of Fig. 2b represents a one-dimensional analogue of the electronic lattice we are familiar with from semiconductor physics. Under Bragg condition an infinite lattice displays total reflection. Such photonic mirrors are frequently used in photonics and semiconductor lasers. They are specified by a higher reflectivity than metallic mirrors.

An incoming electromagnetic wave $\mathbf{E}^{\text{in}}$ will be partially reflected and partially transmitted by a finite barrier. The ratio between reflected and incoming wave defines the reflection coefficient $r = E_n^{\text{reflected}}/E_n^{\text{in}}$ and the ratio between the transmitted wave and the incoming wave the transmission coefficient $t = E_n^{\text{trans}}/E_n^{\text{in}}$ where E_n denotes the normal component of the vector $\mathbf{E}$. In general, the transmission and reflection coefficients t and r are complex

$$t = \sqrt{T}e^{i\varphi_t} \tag{12}$$

$$r = \sqrt{R}e^{i\varphi_r}, \tag{13}$$

where the transmissivity $T = |t|^2$ and the reflectivity $R = |r|^2$. They are related due to conservation of energy as

$$T + R = 1\,. \tag{14}$$

The one-dimensional lattice introduced and studied here is built up by layers with a periodic alteration of the refractive index. The elementary cell is given by the two quarter wave length layers of thicknesses d_1 and d_2

$$n_1 d_1 = n_2 d_2 = \lambda_0/4 \tag{15}$$

$$\omega_0 = 2\pi c/\lambda_0, \tag{16}$$

where ω_0 corresponds to the mid-gap angular frequency of such an arrangement. The mid-gap frequency is given by the resonance condition (15) and is displayed as transmission minimum in Fig. 3. (Fig. 3 shows the transmission gaps of two structures and Fig. 9b displays the frequency spectrum of a signal displaced in the middle of the forbidden frequency gap.) Next I will calculate the transmission function for the quarter-wavelength unit cell. The transmission coefficient as defined above can be given by the complex number

$$t = \sqrt{T}e^{i\varphi_t} \tag{17}$$

The transmission for the quarter wave stack is

$$t = \frac{T_{12}e^{i(p+q)}}{1 - R_{12}e^{2iq}}, \tag{18}$$

where

$$T_{12} = t_{12}t_{21} = \frac{4n_1n_2}{(n_1 + n_2)^2}, \tag{19}$$

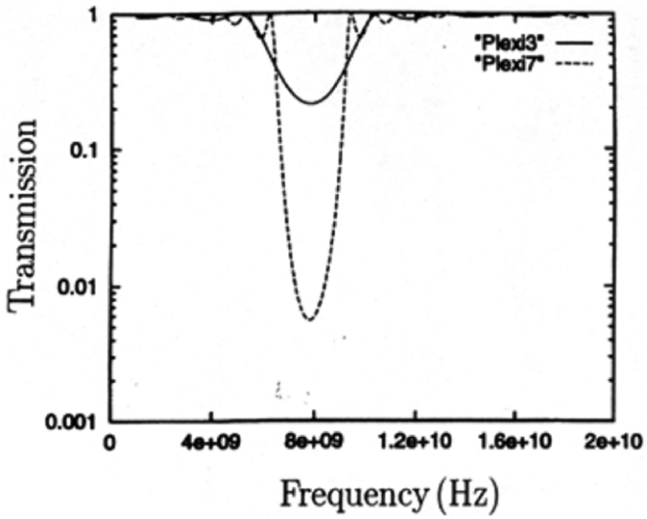


Fig. 3. Transmissivity T as a function of frequency of periodic dielectric quarter-wavelengths structures with 7 and with 3 perspex layers, respectively. The interference pattern shows small minima due to multiple layer destructive interference

and

$$R_{12} = r_{12}^2 = \left(\frac{n_1 - n_2}{n_1 + n_2}\right)^2 , \tag{20}$$

are the double-transmission and reflection factors. Here, $p = n_1 d_1 \omega/c$ and $q = n_2 d_2 \omega/c$, where 15 holds in addition for a quarter wave stack. After extracting the real and the imaginary parts from the quarter cell transmission coefficient, $t^{\lambda/4}$, we have unit-cell expressions for $x^{\lambda/4}$ and $y^{\lambda/4}$,

$$x^{\lambda/4} = T_{12} \frac{\cos(\pi\omega') - R_{12}}{1 - 2R_{12}\cos(\pi\omega') + R_{12}^2} , \tag{21}$$

$$y^{\lambda/4} = T_{12} \frac{\sin(\pi\omega')}{1 - 2R_{12}\cos(\pi\omega') + R_{12}^2} , \tag{22}$$

$$T^{\lambda/4} = \frac{T_{12}^2}{1 - 2R_{12}cos(\pi\omega') + R_{12}^2} , \tag{23}$$

where $\omega' = \omega/\omega_0$.

The extension of the relations for an $N = 1$ stack to N an arbitrary number of stacks is presented in Refs. [13, 14], for example.

Numerical data for 3 and 7 stacks are displayed in Figs. 3; 4; 5. Figs. 3–7 illustrate the transmission, the phase shift, and the group velocity as a function of frequency of a photonic lattice in the microwave frequency regime. The data are in agreement with the experiments [3, 5]. The lattice has either 7 or 3 quarter-wavelength perspex layers, which are separated by quarter-wavelength air distances (in this example the refractive indices are $n_1 = 1.6$ and $n_2 = 1.0$)

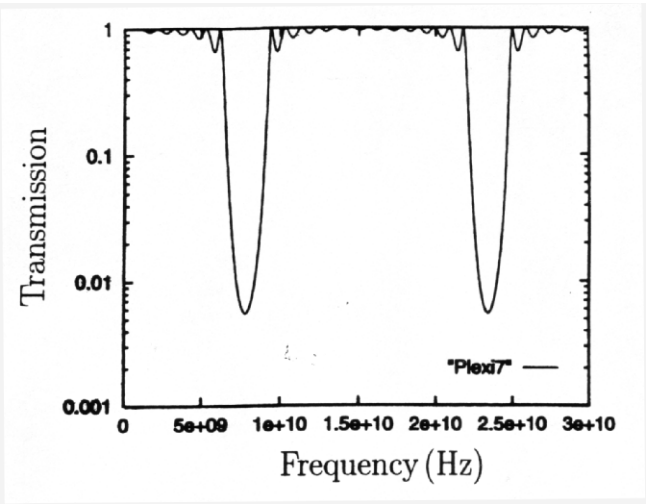


Fig. 4. The transmissivity displays two forbidden band gaps in this frequency range between 0 and 30 GHz. Gaps appear periodically in frequency, in this graph by about 8 and 24 GHz. The data shown is valid for $N = 7$ layers

Figure 5 displays for both structures a reduced phase derivative in the forbidden evanescent frequency regime. *This very derivative equals the above mentioned scattering phase shift and equivalently the scattering time at the barrier front boundary.* The phase time approach is made plausible in Sect. 5.2

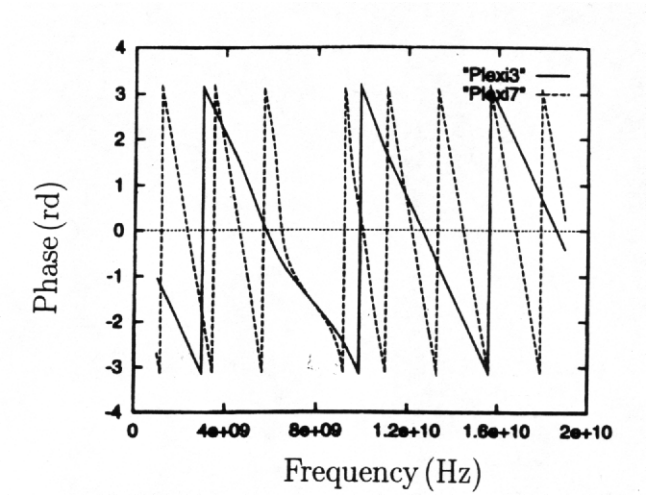


Fig. 5. Phase vs frequency of the lattice. The small phase shift in the evanescent regime of the lattice is due to the phase shift at the barrier front boundary. Inside the barrier the phase shift is zero in consequence of the imaginary wave number

Calculated and measured spectra of the transmission of an infrared periodic dielectric hetero-structure are shown in Fig. 9b.

3.3 Frustrated Total Internal Reflection: The Double Prisms

Double prisms are subject of research since Newton. He already conjectured the Goos-Hänchen shift. This non-specular reflection effect was measured only in 1947. A hundred years ago J. C. Bose studied the transmission of radio wave intensity, i.e. tunneling depending on the gap length [20]. The following dispersion relation describes the frustrated total internal reflection (FTIR) of double prisms. In the case of double prisms the total reflection is called frustrated since a small amount of the incident beam is tunneling into the second prism as sketched in Fig. 2c. In the case of FTIR the imaginary wave number κ in the barrier region and the tunneled electric field E_t measurable outside the barrier are given by the relations [6]

$$\kappa = \frac{\omega}{c}\sqrt{\frac{n_1^2}{n_2^2}\sin^2\theta - 1}\,, \tag{24}$$

$$E_t = E_0(x)e^{i\omega t - \kappa x}, \tag{25}$$

where θ is the angle of the incident beam (larger than the angle of total reflection), $E(x = 0)$ the amplitude of the electric field at the barrier front, n_1 and n_2 are the refractive indexes, and $(n_1/n_2)\sin\theta > 1$ holds in the case of total reflection. ω is the angular frequency, t the time, x the distance of the prisms, and κ the imaginary wave number of the tunneling mode. *Incidentally, n_1 and n_2 do not represent the effective refractive index of the evanescent mode traversing the gap between the prisms, the latter being imaginary.*

Equation (24) is derived from reflection of a beam at the surface of a medium with refractive index n_2. The incident beam comes from a material with a real index n_1 greater than n_2 under the angle θ_i. Snell's law says that

$$n_1 \sin\theta_i = n_2 \sin\theta_t. \tag{26}$$

The angle θ_t of the transmitted wave becomes 90^0 when the incident angle θ_i is equal to the critical angle θ_c given by

$$\frac{n_1}{n_2}\sin\theta_c = 1. \tag{27}$$

The magnitudes of the wave vectors, that is, of the incident wave number $\boldsymbol{k}$, of the reflected wave vector $\boldsymbol{k}_r$, of the transmitted wave vector $\boldsymbol{k}_t$, the wave vector parallel $\boldsymbol{k}_\parallel$ and perpendicular $\boldsymbol{k}_\perp$ to the boundary follows from the boundary conditions at the interface. They are given by

$$k^2 = n^2\omega^2/c^2 \tag{28}$$
$$k_t^2/n_2^2 = k_r^2/n_1^2 = k^2/n_1^2 \tag{29}$$
$$k_{t,\parallel} = k_{r,\parallel} = k_\parallel \tag{30}$$
$$k_{r,\perp}^2 + k_{r,\parallel}^2 = k_\perp^2 + k_\parallel^2 \tag{31}$$
$$k_{r,\perp}^2 = k_\perp^2 \tag{32}$$
$$k_{r,\perp} = -k_\perp\,. \tag{33}$$

With $k_t^2 = k_{t,\parallel}^2 + k_{t,\perp}^2$ we can find

$$k_{t,\perp}^2 = k_t^2 - k_{t,\parallel}^2 \tag{34}$$
$$k_{t,\perp}^2 = \frac{n_2^2}{n_1^2}k^2 - k_{r,\perp} \tag{35}$$
$$k_{t,\perp}^2 = \omega^2\frac{n_2^2}{c^2}\left(1 - \frac{n_1^2}{n_2^2}\sin^2\theta_i\right)\,, \tag{36}$$

with $k_\parallel = k\sin\theta_i$. The last equation equals the dispersion relation (24) in the case of FTIR given above.

4 Evanescent Modes Are not Observable

Remarkable, evanescent modes like tunneling particles are not observable inside a barrier [21–23]. For instance, evanescent modes don't interact with an antenna as long as the system is not perturbed, i.e. the evanescent mode is not transformed back into a propagating electromagnetic wave. Evanescent modes like tunneling particles display some outstanding properties:

(1) The electric energy density u of the evanescent electric field E with its imaginary refractive index is negative:

$$u = \frac{1}{2}\epsilon E^2 \ < 0 \tag{37}$$
$$\epsilon \ = n^2 \ < 0. \tag{38}$$

In the case of particle tunneling we have a negative total kinetic W energy:

$$W = W_{\text{kin}} - U_0 \ < 0, \tag{39}$$

where W_{kin} and U_0 are the kinetic energy and the potential barrier height, respectively. Equation (37) is the quantum electrodynamic basis for the existence of evanescent waves. It has been shown by Ali [24] that virtual photons are those modes which do not satisfy the Einstein relation $W^2 \neq (\hbar k)^2c^2$.

(2) An evanescent field does not interact with real fields due to the imaginary wave number resulting in a refractive index mismatch. Fields can only transmit energy if for the reflection $R < 1$ holds. If n_1 represents the imaginary refractive index of an evanescent region and n_2 represents the refractive index of the dielectric medium then the square of the absolute value gives

$$R = \mid r \mid^2 = \frac{\mid n_2 - n_1 \mid^2}{\mid n_2 + n_1 \mid^2} \tag{40}$$

equals 1 and total reflection takes place.
In order to observe a particle in the barrier it must be localized within a distance of order $\Delta x \approx 1/\kappa$. Hence, its momentum Δp must be uncertain by

$$\Delta p > \hbar/\Delta x \approx \sqrt{2m(U_0 - W_{\text{kin}})} \tag{41}$$

The particle of energy W_{kin} can thus be located in the nonclassical region only if it is given an energy $U_0 - W_{\text{kin}}$, sufficient to raise it into the classically allowed region [22, 23].

(3) The quantization of evanescent modes by Carniglia and Mandel has shown that the locality condition is not fulfilled [25]. They figured out that the commutator of the field operator does not vanish for space-like separated points. This important point is discussed for EPR-correlations by Mittelstaedt [26].

5 Velocities, Delay Times, and Signals

We shortly define and discuss the velocities which can be associated to the various propagation phenomena of waves. *Remember, we are exclusively interested in the propagation of a cause, which is given by the signal velocity only.* We shall become aware that evanescent modes as well as tunneling wave packets are traveling independent of time. First the different quantities are made plausible by the sketch of two traveling wave packets displayed in Fig. 6. The wave packets are representative for voltage pulses of digital signals. The voltage oscillates with a frequency $\omega_0 \pm \Delta\omega$. The pulses begin and end gradually with time. In consequence a physical signal has no well defined front and front velocity. A well defined front and tail of a signal would presuppose an infinite frequency band width and then an infinite energy in consequence of $\hbar\omega$.

The phase velocity is given by the motion of a point stuck to the oscillations. The group velocity is given by the speed of the maximum of the packet, i.e. by the maximum of the pulse. The two velocities are equal in vacuum, but they may differ if traveling through interacting matter like glass or along a wave guide, for instance. The signal velocity is given by the speed of the envelope in order to measure the signal, which is in this example the indicated time duration at half pulse peak of the two pulses, so called half width. As seen by inspection of this figure, the signal and then the half width does not depend on its magnitude. In dispersive media with $n = f(\omega)$ group and signal velocities can be strongly different and the signal may be reshaped and lost its information, i.e. the cause. Such an example of pulse reshaping is displayed in Fig. 13. Essentially, here we are interested in the problem of causality, in cause and subsequent effect. A signal and then an effect can only be detected by its energy. In this respect signal and energy velocities are equal.

The notions on velocity and wave propagation are presented in many text books see Refs. [8, 11, 12], for instance.

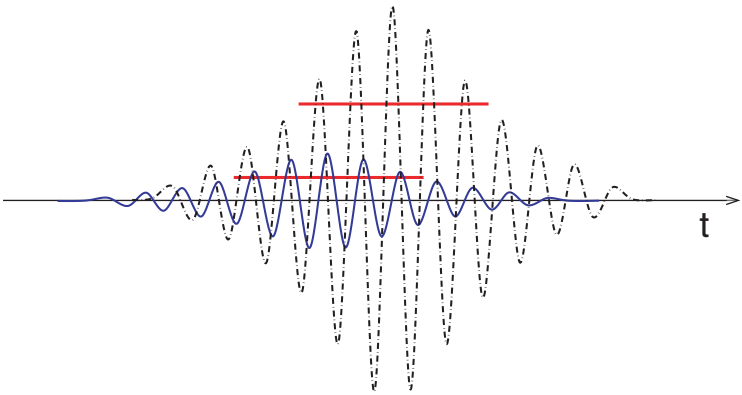


Fig. 6. Sketch of the oscillations of two wave packets (i.e. pulses) vs time. The larger packet has traveled slower than the attenuated one. The horizontal bars indicate the half width of the packets, which do not depend on the packet's magnitude. The figure illustrates the gradually beginning of the packets. The forward tail of the smooth envelope may be described by the relation $(1 - \exp(-t/\tau))\sin(\omega t)$ for instance, where τ is a time constant

5.1 Phase Velocity

Generally the phase velocity can be described by $\phi(\mathbf{x}, t) = Ae^{iS(\mathbf{x},t)}$, where $S(\mathbf{x}, t)$ is the phase of the wave and A its constant amplitude. The phase velocity is the speed related to the trajectory defined by the condition that the phase $S(\mathbf{x}, t)$ is constant, $S(\mathbf{x}, t) = const$. This condition relates $\mathbf{x}$ and t. From

$$0 = dS(\mathbf{x}, t) = \boldsymbol{\nabla} S \cdot d\mathbf{x} + \frac{\partial S}{\partial t} dt \tag{42}$$

and the definitions of the wave vector and the angular velocity

$$\mathbf{k} := -\boldsymbol{\nabla} S\,, \qquad \omega := \frac{\partial S}{\partial t} \tag{43}$$

we immediately obtain $\mathbf{k} \cdot \mathbf{v} = \omega$, where $\mathbf{v}_{\text{phase}} = d\mathbf{x}/dt$. With the dispersion relation $\omega = ck/n$ we obtain

$$\mathbf{k} \cdot \mathbf{v}_{\text{phase}} = \frac{c}{n(\omega)} k \tag{44}$$

or

$$v_{\text{phase,k}} = \frac{c}{n(\omega)}\,, \tag{45}$$

where $v_{\text{phase,k}}$ is the component of $\mathbf{v}_{\text{phase}}$ parallel to $\mathbf{k}$.

5.2 Group Velocity

The group velocity describes the velocity of the modulation of a harmonic traveling wave or of the peak of a wave packet. It is defined by

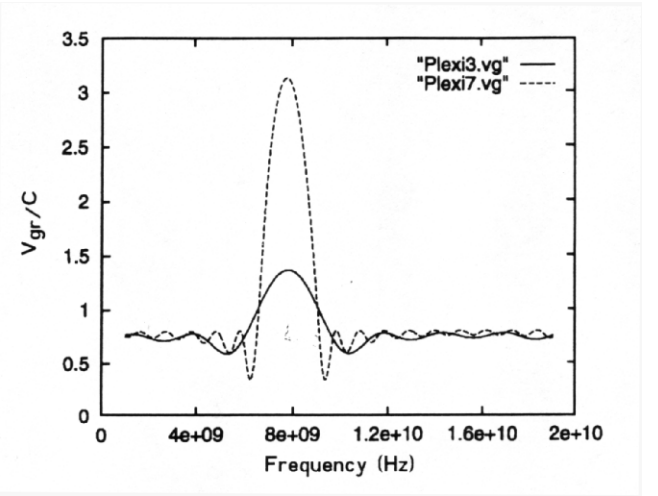


Fig. 7. Calculated group velocity vs frequency for two multiple layer structures as follows from (46) and Fig. 5

$$\mathbf{v}_{\text{group}} = \boldsymbol{\nabla}_k \omega(\mathbf{k})\,, \tag{46}$$

and represents the first term of a Taylor series of the modulation velocity. In vacuum, the group velocity equals c. See Fig. 7 for the group velocity in the case of a photonic lattice.

The group velocity can be rewritten as

$$v_{\text{group}} = \frac{d\omega}{dk} = \frac{c}{n(\omega) + \omega dn(\omega)/d\omega} \tag{47}$$

The last relation is interesting, as it elucidates the difference between the phase and the group velocity. It is the second term of the denominator, which distinguish the group from the phase velocity. For instance, in glass the group is about 2 % slower than the phase in the visible range of the spectrum.

We also have

$$v_{\text{group}} = \frac{x}{t_{\text{group}}} = \frac{x}{\partial S/\partial\omega}, \tag{48}$$

where $t_{\text{group}} = \partial S/\partial\omega$ is the group time delay or phase time. The phase shift is given by $\partial S = x\partial k$ in the region x considered. The group time delay represents the time delay of a maximum for traversing a distance as displayed in Fig. 13 for a strong dispersion. The case of a negligible dispersion shown in Figs. 1; 9. In the latter case the group time delay represents the time delay of a signal and of the energy.

5.3 Signal Velocity

A signal carries information which is a defined cause with a subsequent effect. For a simple example see digital signals shown in Fig. 8. Digital signals are given

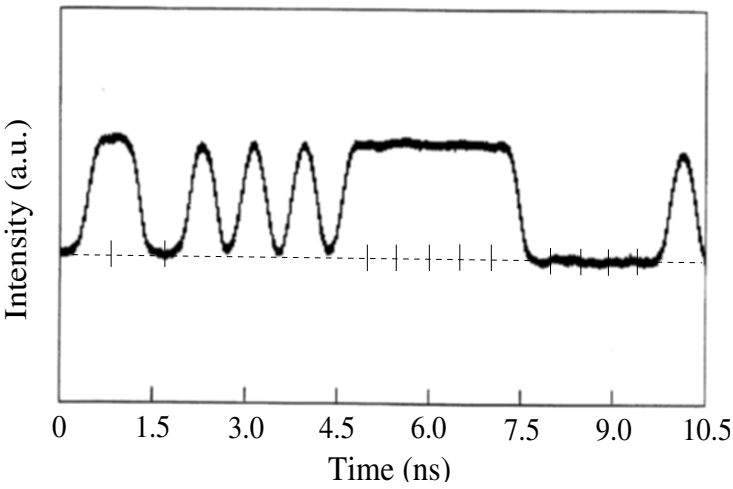


Fig. 8. Signals: Measured signal intensity in arbitrary units. The half width in units of 0.2 ns corresponds to the number of bits. From left to right: 1,1,0,0,1,0,1,0,1,0,1,1,1,1,1.... The infrared carrier frequency of the infrared signal is $2 \cdot 10^{14}$ Hz (wavelength 1.5 µm). The frequency-band-width of the signal is about $2 \cdot 10^{9}$ Hz corresponding to a relative frequency-band-width of 10^{-5} [17]

by their half width (the half width is the time span between the half power points, see e.g. Figs. 1; 6; 8; 9).

In general signals are characterized by their envelope, whether we are transmitting Morse signals, a word or a melody, always the complete envelope has to be measured, see [12], for instance. Therefore, the signal velocity in vacuum is identical to the group velocity in the case of negligible dispersion:

$$\mathbf{v}_{\text{signal}} \equiv \mathbf{v}_{\text{group}} \qquad (\text{in vacuum})\,. \tag{49}$$

Delay times and velocities are quantities depending only on the real part of the refractive index n' and on the derivative of the phase $S = 2\pi k_0 n' x$. In the case of evanescent modes or tunneling with a purely imaginary refractive index n'' the phase S is constant. *Thus according to (48) the group time delay becomes* $\rightarrow$ *0 and the group velocity* $\rightarrow \infty$. There is measured a phase shift and thus a short delay time corresponding to about one oscillation time of the signal in tunneling. This scattering time occurs at the front boundary and not inside the evanescent region nor inside a potential barrier. In the case of microwave pulses this time is about 100 ps and in the infrared case of glass fiber communication about 5 fs [13, 15], see also the data displayed in Fig. 5 and its interpretation. As this scattering time is independent of barrier length for opaque barriers with $\kappa x \geq 1$ (the so called Hartman effect) the effective group velocity (48) increases with barrier length [16]. This behavior is illustrated below in Figs. 5; 7.

The lack of phase shift means a zero-time barrier traversal of evanescent modes according to the phase time approach of (48). Actually this zero time

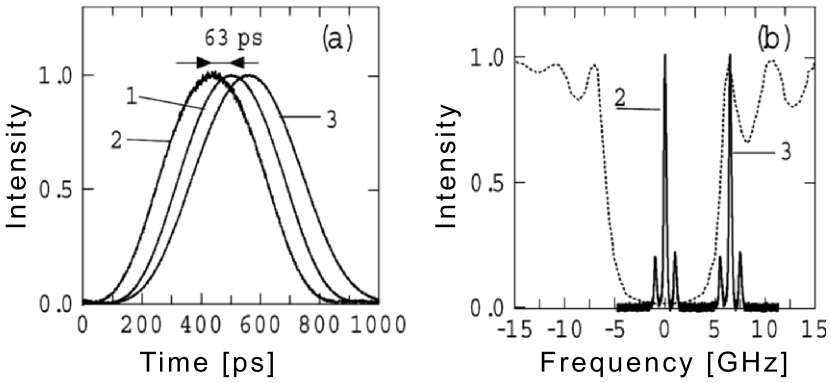


Fig. 9. (**a**) Measured propagation time of three digital signals and spectrum of the photonic lattice transmission [18]. Pulse trace 1 was recorded in vacuum. Pulse 2 traversed a photonic lattice in the center of the frequency band gap (see spectrum in part (**b**) of the figure), and pulse 3 was recorded for the pulse traveling through the fiber outside the forbidden band gap. The tunneling barrier was a photonic lattice of a quarter wavelength periodic dielectric hetero-structure fiber. The frequency zero point in part (b) corresponds to the infrared signal carrier frequency of $2 \cdot 10^{14}$ Hz and to the mid frequency of the forbidden frequency gap of the lattice

was measured in different experiments and the observed short barrier traversal time τ arises as scattering time at the barrier front boundary only [13, 15].

Infrared digital signals used in modern communication systems are displayed in Fig. 8. Such a single digit is tunneled and its velocity is compared with a vacuum and with a fiber traveled signal as shown in Fig. 9. Here Longhi et al. [18, 19] performed superluminal tunneling of infrared pulses over distances up to 50 mm at an infrared signal wavelength of 1.5 μm ($2 \cdot 10^{14}$ Hz). Results are presented in Fig. 9(Curve 1 luminal signal, 2 superluminal, 3 subluminal velocity). The frequency band width is $< 2 \cdot 10^9$ Hz. The measured velocity was $2\,c$ and the transmissivity of the barrier was 1.5%. The narrow band width of the signal is displayed in Fig. 9b. The superluminal signal pulse trace (2) has only evanescent frequency components around the mid frequency of the forbidden frequency gap of the photonic barrier.

5.4 The Front Velocity

As mentioned above the front velocity is an idealized notion and, thus, has no precise physical meaning. It is presupposing an infinite frequency band width of a signal. Its definition is given by

$$v_{\text{front}} = \lim_{\omega \to \infty} \frac{\omega}{k}\,. \tag{50}$$

Mathematically a discontinuity of the field under consideration or of one of its derivatives will propagate with the front velocity. The normal $(\omega, \mathbf{k})$ of the 3-dimensional hypersurface in 4-dimensional space-time where such discontinuities

may occur is defined by the characteristic equation

$$\omega^2 - \frac{c^2}{n^2}\mathbf{k}^2 = 0\,. \tag{51}$$

The velocity in configuration space related to the propagation of these singularities is then given by

$$\mathbf{v}_{\text{front}} = \boldsymbol{\nabla}_k \omega(\mathbf{k}) = \frac{c}{n}\widehat{\mathbf{k}}\,. \tag{52}$$

Therefore $v_{\text{front}} = c/n$. In vacuum, the front propagates with the velocity of light c. Though being a clear mathematical concept, it can be realized in physics only approximatively: Since a discontinuity is described by a Heaviside function (a function $H(x)$ which is zero for $x < 0$ and 1 for $x \geq 0$), the support of its Fourier transform is unbounded, that is, one needs waves with frequencies up to infinity in order to prepare a jump in the propagating field. This needs infinite energy which of course is not available. Therefore, since in reality only a finite range of frequency is available (frequency band limited signals), a jump in the propagating field cannot be created. However, there is no known *fundamental* limit for an upper energy bound (except perhaps the energy available in the universe). Therefore the front velocity is operationally not well defined and has no precise *physical* meaning [8, 12].

5.5 The Energy Velocity

Usually text books present the energy velocity by the relation ship

$$\mathbf{v}_{\text{energy}} = \mathbf{P}/u, \tag{53}$$

where

$$\mathbf{P} = \epsilon_0 c^2 \mathbf{E} \times \mathbf{B} \tag{54}$$

is the Poynting vector, E the electric field, B the magnetic field, and u is the energy density. The Poynting vector represents the energy flux and subtracts transmitted and reflected flux, whereas the scalar energy density adds both transmitted and reflected energy densities. This approach is then only correct in the case of no reflection and can not be applied for evanescent modes or tunneling, see e.g. [3]. The attenuation of evanescent modes is not due to dissipation but due to reflection. Equation (53) even can not be used to calculate the energy velocity in an open coaxial transmission line. Due to the impedance mismatch at the open end there takes place a strong reflection and (53) gives a too slow energy velocity for the energy loss at the end of the coaxial transmission line.

As already mentioned, we are interested in the *effect of a cause*. From this condition we can conclude that the energy velocity equals the signal velocity: A signal is received by an inelastic detection process. So it is the signal's energy which result in an defined effect.

6 Partial Reflection: An Experimental Method to Demonstrate Superluminal Signal Velocity of Evanescent Modes

Superluminal signal velocities were observed in different transmission and partial reflection experiments [5, 18–20]. The short tunneling and reflection times are equal. The result shows that the measured short time is spent at the barrier entrance. Inside a barrier the wave packet spends zero time. Transmission and reflection times are independent of barrier length as was calculated with the Schrödinger equation by Hartman and measured later [5, 13, 15, 27, 28]. This Hartman effect holds for opaque barriers with $\kappa x \geq 1$. The result demonstrate the nonlocal properties of evanescent modes and of the tunneling process as was shown by Carniglia and Mandel for instance [25].

A smart experimental set-up to measure both the transmission and the reflection times at the same time is sketched in Fig. 10. The distances of the reflected and of the transmitted beams differ only by the gap between the two prisms, i.e. the evanescent region (tunneling distance). *It was measured the same traveling time for both the reflected and the transmitted signals, obviously tunneling took place in zero-time [20].* The result was revisited by Stahlhofen [29] and was conjectured by quantum mechanical calculations for electron tunneling by Hartman and later by Low and Mende [27, 30]. The latter authors write that traversing a barrier appears to do so in *zero time*.

The reflection by a photonic lattice at a frequency of its forbidden band gap (see e.g. Fig. 9b) is measured and compared with the time crossing the same distance between two metallic mirrors. One mirror is positioned at the barrier

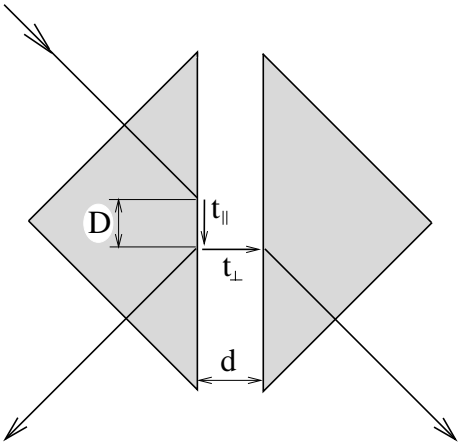


Fig. 10. Symmetrical FTIR set-up to measure both the reflection and the transmission time of a double prism, where $t_\perp$ is the time traversing the gap d and $t_\parallel$ is the time spent for traveling along the boundary of the first prism. The latter represents the time of the Goos–Hänchen shift [20]. The measured reflection time equals the transmission time resulting in a zero tunneling time $t_\perp = 0$

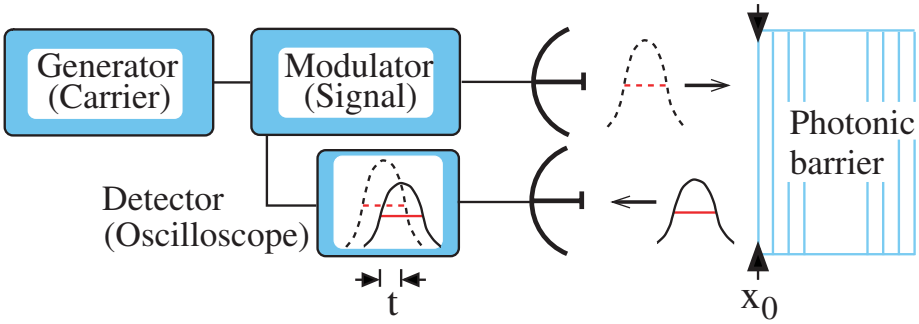


Fig. 11. Set-up to measure the time dependence of partial reflection at a photonic barrier with a digital pulse. The parabolic antenna on top of the illustration transmit digital pulses toward the barrier, the second one below receives the reflected signal. The time delay is measured with the oscilloscope

front side and the other one at the barrier back side. The set-up and the results are shown in Figs. 11, 12. The measured reflected time equals the time measured for the mirror's front position neglecting the mentioned short interaction time at the barrier front. The amazing result is that barrier height and barrier length

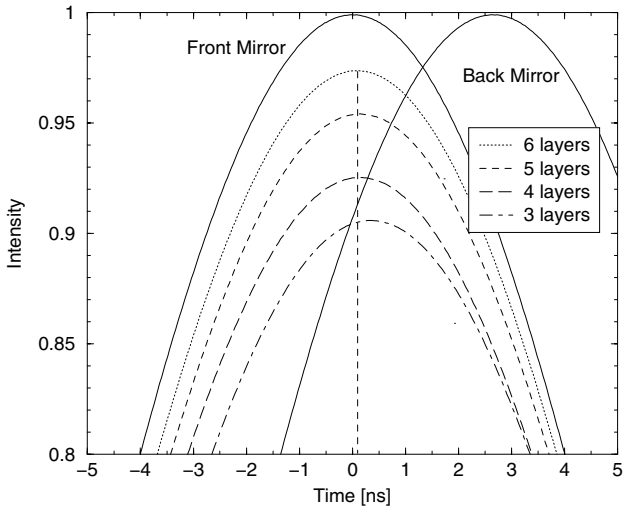


Fig. 12. Measured partial reflected microwave pulses vs time. Parameter is the barrier composition as illustrated in Fig. 11. The signal reflections from metal mirrors either substituting the barrier's front or back positions are displayed [28]. In this experiment the wavelength has been 3.28 cm and the barrier length was 40 cm. The number of lattice layers was reduced from 6 to 3 inside the resonant lattice structure illustrated in Fig. 11

are instantaneously displayed in the reflected signals as seen from inspection of Fig. 12.

The performance demonstrates that the reflected signal carries the information about barrier height and barrier length at the same time when the signal is reflected by the front mirror. The reflection time is independent of barrier length, the field spreading inside the barrier is instantaneous. The reflection amplitude decreases with decreasing barrier length but the reflection time is constant in the case of opaque barriers with $\kappa x \geq 1$.

7 Evanescent Modes a Near Field Phenomenon

According to many text books and review articles, superluminal signal velocities are violating Einstein causality, implying that cause and effect can be interchanged and that time machines known from science fiction can be designed [31–33]. Actually, it can be shown for frequency band *unlimited groups* that the front travels always at a velocity $\leq c$, and only the peak of the pulse has traveled with a superluminal velocity. As mentioned above such calculations were carried out by several authors, for example [34–36]. In this case the tunneled pulse is reshaped and its front has propagated at luminal velocity.

However, this approach does not describe physical signals as those signals displayed for instance in Figs. 1; 6; 8; 9. In this case the signal has gradually formed a front tail. A pulse reshaping did not happen and the envelope of the signal traveled at a superluminal velocity.

Pulse reshaping of a frequency band unlimited signal is displayed in Fig. 13. The half width of this artificial pulse with a discontinuous front step is significantly reduced compared with the original signal and only the pulse peak has traversed the barrier at superluminal velocity [34].

Frequently it is claimed that a tunneled small signal would not cross the front tail of the original signal, see for instance [34–36]. The argument is taken to prove that superluminal signal velocities are not allowed and do not occur. The frequency band limited digital signals presented in Figs. 1; 9 are crossing each other. This result is in consequence of the fact that these superluminal pulses contain only evanescent frequency components.

A physical signal can not be described by a Gauss function having an infinite frequency band. For a physical signal the relation [11, 37]

$$\Delta\nu \cdot \Delta t \geq 1, \tag{55}$$

holds, whith both $\Delta\nu$ and $\Delta t \ll \infty$. Such a pulse of field oscillations is sketched in Fig. 6. Actually, relation (55) is proportional to the information content of a signal as was shown by Shannon [37]. According to Fourier transform such a physical signal with both limited frequency band and time duration is not causal [12, 38]. On the other hand it is obvious that a physical signal has to be frequency band limited. Signals start gradually within a time span given by its frequency band width [8, 12].

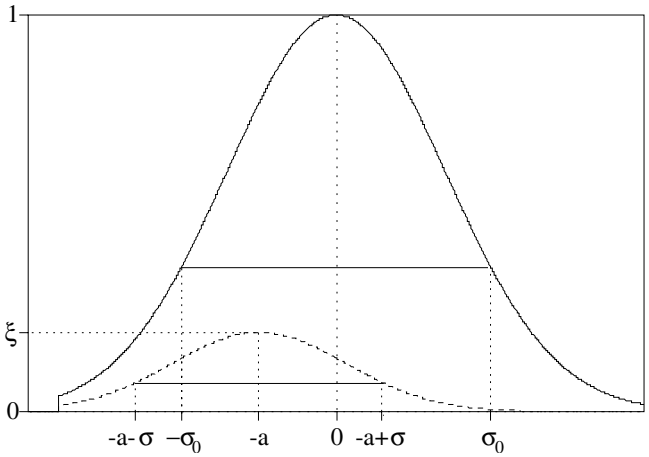


Fig. 13. Comparison of calculated intensity vs time of an airborne pulse (*solid line*) and the same tunneled pulse (*dotted line*) [34]. Both signals have a sharp step at the front and thus an infinite frequency bandwidth. The tunneled signal is reshaped and attenuated. Its maximum has traveled at superluminal velocity. Both fronts have traversed the same distance with speed c, ξ is the maximum of the tunneled pulse, a is the shift of the maximum, σ is the halfwidth of the tunneled signal, and σ_0 is the halfwidth of the airborn signal [34]. The halfwidths $\sigma \gg \sigma_0$ holds, i.e. the digital information is strongly reshaped

As the Gauss function does not describe a physical signal, mathematicians and engineers have developed a number of so called window functions [39]. They are limited in both frequency and time but can be quasi causal transformed from time to frequency domain and vice versa.

For example, physical digital signals are well described by the Kaiser-Bessel function for instance. This function is used in network analyzers describing the intensity vs time as well as the frequency band of physical signals. This function allows even a causal Fourier transform from time domain to frequency domain down to intensities at which the Johnson noise limits detectors finally, see (57, 58).

In Figs. 14 and 15 the Kaiser-Bessel function is plotted as a function of intensity I(t) vs time. The curves can be scaled to the data of the experiments displayed in Figs. 1 and 9.

The Kaiser–Bessel function – often called Kaiser–Bessel window as time duration and frequency spectrum are limited – is given

$$I(t) = \frac{I_0\left(\pi \Delta t \Delta\nu (1 - \sqrt{\frac{t}{\Delta t/2}})\right)}{I_0 \pi \Delta t \Delta\nu}, \tag{56}$$

where I_0, and $\pi \Delta t \Delta\nu$ are the zero-order modified Bessel function of the first kind, and the time-bandwidth product, respectively. $0 \leq |t| \leq \Delta t/2$, represents the investigated time interval.

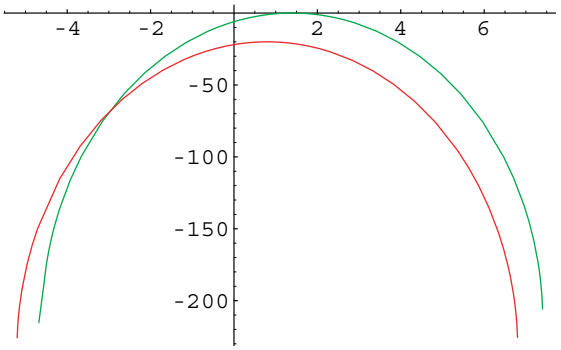


Fig. 14. Calculated pulse intensity of the Kaiser–Bessel function vs time in a.u.. The data can be scaled to the measured pulses displayed in Figs. 1; 8; 9. In the graph the tunneled signal is attenuated by −20 dB

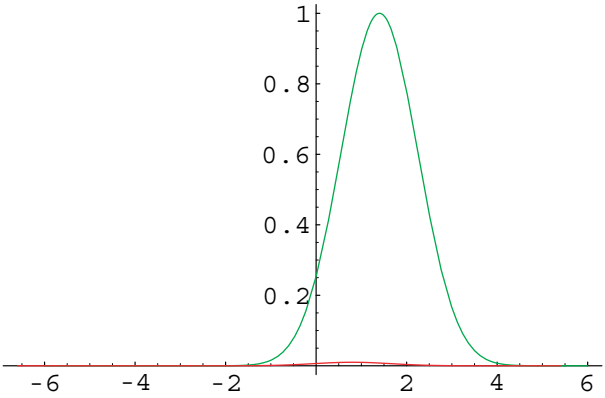


Fig. 15. The same data as shown in Fig. 14 in a semi-logarithmic plot. The ordinate is scaled in dB and the abscissa in a.u.

A signal can be detected only if its power is above the Johnson noise P_{JN}. The thermal noise was observed and measured by Johnson in 1928 and is theoretically elaborated by the Nyquist Theorem, see for instance [40]. The theorem is of great importance in experimental physics and electronics. It is concerned with the spontaneous thermal fluctuations of voltage across an electric circuit element. The theorem gives a quantitative expression for thermal noise power generated by a resistor in thermal equilibrium:

$$P_{\mathrm{JN}} = kT\Delta f, \tag{57}$$

This relation yields a classical estimate for the near field extension of evanescent modes. The power $P(x)$ of a signal, i.e. of a defined effect has to be detected. Then superluminal signal propagation is limited by the relationship, which gives the minimum tunneled signal power:

$$P(x) = P_0 e^{-2\kappa x} \geq kT\Delta f, \tag{58}$$

where P_0 is the incident power of the evanescent mode, κ is the imaginary wave number of the evanescent mode, x the length of the evanescent region, k the Boltzmann constant, T the temperature, and Δf the frequency range of the signal. For example an infrared signal source of 1 mW power, a carrier frequency of $2 \cdot 10^{14}$ Hz (1.5 µm wavelength), and an imaginary wave number in the barrier $\kappa = 115\,\mathrm{m}^{-1}$ at a temperature of $T = 300\,\mathrm{K}$. Thus the Johnson noise with $\approx 1\,\mu\mathrm{W}$ limits a detectable near field up to 0.03 m, corresponding to about 20000 wavelengths of this infrared digital signal and this special photonic barrier. In the above introduced microwave experiments the near field was limited to less than a hundred wave lengths.

8 Superluminal Signals Do not Violate Primitive Causality

Does the measured superluminal signal velocity violate the principle of causality? The line of arguments showing how to manipulate the past in the case of superluminal signal velocities is illustrated in Fig. 16. There are displayed two frames of reference. In the first one lottery numbers are presented as points on the time coordinate with zero time duration. At $t = 0$ the counters are closed. Mary (A) sends the lottery numbers to her girl friend Susan (B) with a signal velocity of 4; c. Susan, moving in the second inertial system at a relative speed of 0.75 c, sends the numbers back at a speed of 2 c, to arrive in the first system of Mary at $t = -1$ s, thus in time to deliver the correct lottery numbers before the counters close at $t = 0$.

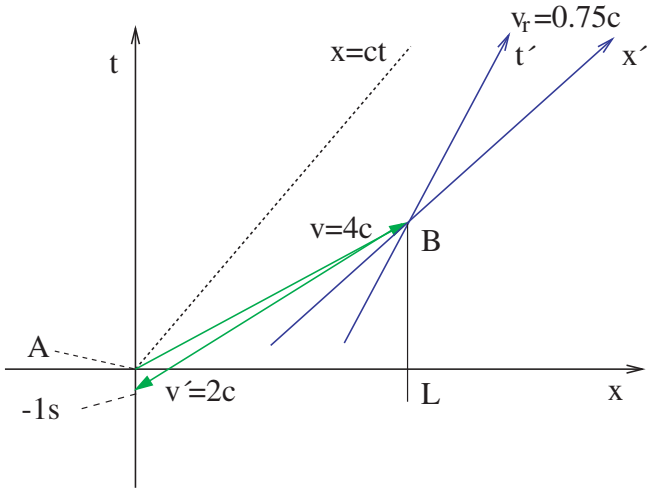


Fig. 16. Coordinates of two inertial observers **A** $(0, 0)$ and **B** with $O(x, t)$ and $O'(x', t')$ moving with a relative velocity of 0.75 c. The distance L between **A** and **B** is 2 000 000 km. **A** makes use of a signal velocity $v_s = 4\ c$ and **B** makes use of $v'_s = 2c$ (in the sketch is $v \equiv v_s$). The numbers in the example are chosen arbitrarily. The signal returns -1 s in the past in **A**

The time shift of a point on the time axis of reference system A into the past is given by the relation, [32, 41],

$$t_A = -\frac{L}{c} \cdot \frac{(v_r - c^2/v_s - c^2/v'_s + c^2 v_r/v_s v'_s)}{(c - c v_r/v'_s)}, \tag{59}$$

where L is the transmission length of the signal, v_r is the velocity between the two inertial systems A and B. The condition for the change of chronological order is $t_A < 0$, the time shift between the systems A and B. This interpretation assumes, however, a signal to be of zero time neglecting its temporal width.

Several tunneling experiments have revealed superluminal signal velocity in tunneling photonic barriers [5]. Nevertheless, the principle of causality has not been violated as will be explained in the following.

In the example with the lottery data, the signal was assumed to be a point in space-time. However, a physical signal has a finite duration like the pulses sketched along the time axis in Fig. 17.

The general relationship for the bandwidth-time interval product of a signal, i.e. a packet of oscillations is given by (55). A zero time duration of a signal would require an infinite frequency bandwidth.

Taking into consideration the dispersion of the transmission of tunneling barriers, the frequency band of a signal has to be narrow in order to suppress non superluminal frequency components and thus pulse reshaping.

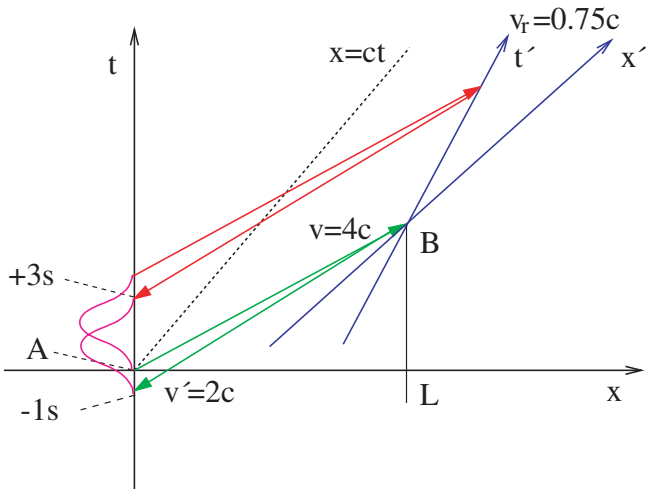


Fig. 17. In contrast to Fig. 16 the pulse-like signal has now a finite duration of 4 s. This data is used for a clear demonstration of the effect. In all superluminal experiments, the signal length is long compared with the measured negative time shift. In this sketch the signal envelope ends in the future with 3 s (in the sketch is $v \equiv v_s$)

Assuming a signal duration of 4 s the complete information is obtained with superluminal signal velocity at 3 s at a positive time as illustrated in Fig. 17.

The compulsory finite duration of all signals is the reason that a superluminal velocity does not violate the principle of causality. A shorter signal with the same information content would have an equivalently broader frequency bandwidth, compare (55). As a consequence, an increase of v_s or v_s' cannot violate the principle of causality.

For instance, the dispersion relation of FTIR (24) elucidate this universal behavior: Assuming a wavelength $\lambda_0 = c/\nu$, a tunneling time $\tau = T = 1/\nu$, and a tunneling gap between the prisms $d = j\,\lambda_0$ $(j = 1, 2, 3, \ldots)$ the superluminal signal velocity is $v_s = j\,c$, (remember the tunneling time is independent of barrier length). However, with increasing v_s the bandwidth $\Delta\nu$ (that is the tolerated imaginary wave number width $\Delta\kappa$) of the signal decreases $\propto 1/d$ in order to guarantee the same amplitude distribution of all frequency components of the signal. In spite of an increasing superluminal signal velocity $v_s \to \infty$ the general causality can not be violated because the signal time duration increases analogously $\Delta t \to \infty$, see (55).

9 Summary

Evanescent modes and tunneling show amazing properties to which we are not used to from classical physics. The tunneling time is short and arises at the barrier front as scattering time. This time equals approximately the reciprocal frequency of the carrier frequency or of the wave packet energy divided by the Planck constant h [13, 15]. Inside a barrier the wave packet does spent zero time [5, 30]. This property results in superluminal signal and energy velocities, as a signal is detected by its energy, i.e. by photons or other field quanta like phonons. The detector receives the tunneled signal earlier than the signal, which traveled the same distance in vacuum as demonstrated in Figs. 1, 9, 12. Evanescent fields like tunneling particles are not observable [22, 23, 25, 42–44].

Another consequence of the frequency band limitation of signals is, if they have only evanescent mode components, as shown for instance in Fig. 9b signal trace (2), they can violate relativistic causality, which claims that signal and energy velocities have to be $\leq c$.

As explained in Sec. evanescent modes and the tunneling process are near field effects. They are roughly limited to the order of the signal length in propagating in vacuum.

In the review on *The quantum mechanical tunnelling time problem - revisited* by Collins et al. [45], the following statement has been made on the much ado about superluminal velocity: *the phase-time-result originally obtained by Wigner and by Hartman is the best expression to use for a wide parameter range of barriers, energies and wave packets.* The experimental results of photonic tunneling have confirmed this statement [5]. In spite of so much arguing about violation of Einstein causality [4, 33, 36, 46, 47], all the properties introduced above are useful for novel devices, for both photonics and electronics [48].

Acknowledgement

I like to thank Claus Lämmerzahl and Alfons Stahlhofen for stimulating and elucidating discussions. The technical advice and support by Horst Aichmann, Astrid Haibel, Harald Spieker, Ralf Vetter, and Igor Drozdov are gratefully acknowledged.

References

1. A. Enders and G. Nimtz, J. Phys. I, France **2**, 1693 (1992)
2. V. Olkhovsky and E. Recami, Phys. Rep. **214**, 339 (1992); E. Recami, Int. J. of Modern Physics, **A 15**, 2793 (2000); V. Olkhovsky, E. Recami, J. Jakiel, Phys. Rep. **398**, 133 (2004).
3. G. Nimtz and W. Heitmann, Progr. Quant. Electr. **21**, 81 (1997).
4. R. Chiao and A. Steinberg, Progress in Optics **XXXVII**, 345 (1997).
5. G. Nimtz, Prog. Quantum Electronics **27**, 417 (2003).
6. R.P. Feynman, R.B. Leighton, and M. Sands, The Feyman Lectures on Physics, **II** 33–12, Addison–Wesley, Reading (1964).
7. G. Nimtz, A. Enders, and H. Spieker, J. Phys. I, France **4**, 565 (1994).
8. L. Brillouin, Wave propagation and group velocity, Academic Press, New York (1960).
9. L. Wang, A. Kuzmich, and A. Dogariu, Nature **406**, 277 (2000).
10. D. Bowmeester, J. Pan, K. Mattle, M. Eibl, H. Weinfurter, and A. Zeilinger, Nature, **390**, 575 (1997).
11. Berkeley physics course, **3**, Chap. 6, McGraw-Hill , New York and London (1968).
12. A. Papoulis, The Fourier Integral And Its Applications, McGraw-Hill, New York, Secs. 7.5 and 7.6 (1962).
13. S. Esposito, Phys. Rev. E **64**, 026609 (2001).
14. J.M. Bendickson, J.P. Dowling, and M. Scalora, Phys. Rev. E **53**, 4107 (1996).
15. A. Haibel and G. Nimtz, Ann. Phys. (Leipzig) **10**, 707 (2001).
16. A. Enders and G. Nimtz, Phys. Rev. E **48**, 632 (1994).
17. E. Desurvivre, Scientific American **266** January, 96(1992).
18. S. Longhi, M. Marano, P. Laporta, and M. Belmonte, Phys. Rev. E **64**, 055602 (2001).
19. S. Longhi, P. Laporta, M. Belmonte, and E. Recami, Phys. Rev. E **65**, 046610 (2002).
20. A. Haibel, G. Nimtz, and A.A. Stahlhofen, Phys. Rev. **E 63**, 047601 (2001).
21. J.P. Fillard, Near field optics and nanoscopy, Worls scientific, Singapore (1997).
22. E. Merzbacher, *Quantum Mechanics*, 2nd edition, John Wiley & Sons, New York (1970).
23. S. Gasiorowicz, Quantum Physics, John Wiley & Sons, New York (1996).
24. S.T. Ali, Phys. Rev. B **7**, 1668 (1973).
25. C.K. Carniglia and L. Mandel, Phys. Rev. D **3**, 280 (1971).
26. P. Mittelstaedt, Ann. Phys. (Leipzig) **7**, 710 (1998).
27. Th. Hartman, J. Appl. Phys. **33**, 3427 (1962).
28. G. Nimtz, A. Haibel and R.-M. Vetter, Phys. Rev. E **66**, 037602 (2002).
29. A.A. Stahlhofen, Phys. Rev. A **62**, 12112 (2000).
30. F.E. Low and P.F. Mende, Ann. Phys. NY **210**, 380 (1991).

31. M. Fayngold, Special Relativity and Motions Faster than Light, Wiley-VCH, Weinheim, 219–223 (2002).
32. R. Sexl und H. Schmidt, Raum-Zeit-Relativität, vieweg studium, Braunschweig (1978).
33. R.U. Sexl and H.K. Urbantke, Relativity, Groups, Particles, Springer, Wien, NewYork (2001).
34. Th. Emig, Phys. Rev. **E 54**, 5780 (1996).
35. G. Diener, Phys. Letters A, **235**, 118 (1997).
36. H. Goenner, Ann. Phys. (Leipzig), **7**, 774 (1998).
37. C.E. Shannon, Bell Sys. Tech. J., **27**, 379, and 623 (1948).
38. D.C. Champeney, Fourier Transforms and their Physical Applications, Academic Press, London and New York (1973).
39. F.J. Harris, Proc. IEEE, **66**, 51 (1978).
40. C. Kittel, Thermal Physics, John Wiley & Sons, New York (1968), pp. 402–405.
41. P. Mittelstaedt, Eur. Phys. J, **B 13**, 353 (2000).
42. G. Nimtz, Gen. Rel. Grav. **31**, 737 (1999); G. Nimtz, Ann. Phys. (Leipzig), **7**, 618 (1998).
43. F. de Fornel, Springer Series in Optical Sciences, **73**, Springer, Berlin (2001).
44. O. Bryngdahl, Progress in Optics, **11**, 167 (1973).
45. S. Collins, D. Lowe and J. Barker, J. Phys. C, **20**, 6213 (1987).
46. M. Büttiker and H. Thomas, Superlattices and Microstructures, **23**, 781 (1998).
47. M. Stenner, D. Gauthier, and M. Neifeld, Nature, **425**, 695 (2003); G. Nimtz and Stenner et al., Nature, **6 May**, (2004).
48. G. Nimtz, IEEE Journal of selected topics in quantum electronics, **9**, 79 (2003).

Lecture Notes in Physics

For information about earlier volumes
please contact your bookseller or Springer
LNP Online archive: springerlink.com

Vol.655: M. Shillor, M. Sofonea, J. J. Telega, Models and Analysis of Quasistatic Contact

Vol.656: K. Scherer, H. Fichtner, B. Heber, U. Mall (Eds.), Space Weather

Vol.657: J. Gemmer, M. Michel, G. Mahler (Eds.), Quantum Thermodynamics

Vol.658: K. Busch, A. Powell, C. Röthig, G. Schön, J. Weissmüller (Eds.), Functional Nanostructures

Vol.659: E. Bick, F. D. Steffen (Eds.), Topology and Geometry in Physics

Vol.660: A. N. Gorban, I. V. Karlin, Invariant Manifolds for Physical and Chemical Kinetics

Vol.661: N. Akhmediev, A. Ankiewicz (Eds.) Dissipative Solitons

Vol.662: U. Carow-Watamura, Y. Maeda, S. Watamura (Eds.), Quantum Field Theory and Noncommutative Geometry

Vol.663: A. Kalloniatis, D. Leinweber, A. Williams (Eds.), Lattice Hadron Physics

Vol.664: R. Wielebinski, R. Beck (Eds.), Cosmic Magnetic Fields

Vol.665: V. Martinez (Ed.), Data Analysis in Cosmology

Vol.666: D. Britz, Digital Simulation in Electrochemistry

Vol.667: W. D. Heiss (Ed.), Quantum Dots: a Doorway to Nanoscale Physics

Vol.668: H. Ocampo, S. Paycha, A. Vargas (Eds.), Geometric and Topological Methods for Quantum Field Theory

Vol.669: G. Amelino-Camelia, J. Kowalski-Glikman (Eds.), Planck Scale Effects in Astrophysics and Cosmology

Vol.670: A. Dinklage, G. Marx, T. Klinger, L. Schweikhard (Eds.), Plasma Physics

Vol.671: J.-R. Chazottes, B. Fernandez (Eds.), Dynamics of Coupled Map Lattices and of Related Spatially Extended Systems

Vol.672: R. Kh. Zeytounian, Topics in Hyposonic Flow Theory

Vol.673: C. Bona, C. Palenzula-Luque, Elements of Numerical Relativity

Vol.674: A. G. Hunt, Percolation Theory for Flow in Porous Media

Vol.675: M. Kröger, Models for Polymeric and Anisotropic Liquids

Vol.676: I. Galanakis, P. H. Dederichs (Eds.), Halfmetallic Alloys

Vol.677: A. Loiseau, P. Launois, P. Petit, S. Roche, J.-P. Salvetat (Eds.), Understanding Carbon Nanotubes

Vol.678: M. Donath, W. Nolting (Eds.), Local-Moment Ferromagnets

Vol.679: A. Das, B. K. Chakrabarti (Eds.), Quantum Annealing and Related Optimization Methods

Vol.680: G. Cuniberti, G. Fagas, K. Richter (Eds.), Introducing Molecular Electronics

Vol.681: A. Llor, Statistical Hydrodynamic Models for Developed Mixing Instability Flows

Vol.682: J. Souchay (Ed.), Dynamics of Extended Celestial Bodies and Rings

Vol.683: R. Dvorak, F. Freistetter, J. Kurths (Eds.), Chaos and Stability in Planetary Systems

Vol.684: J. Dolinšek, M. Vilfan, S. Žumer (Eds.), Novel NMR and EPR Techniques

Vol.685: C. Klein, O. Richter, Ernst Equation and Riemann Surfaces

Vol.686: A. D. Yaghjian, Relativistic Dynamics of a Charged Sphere

Vol.687: J. W. LaBelle, R. A. Treumann (Eds.), Geospace Electromagnetic Waves and Radiation

Vol.688: M. C. Miguel, J. M. Rubi (Eds.), Jamming, Yielding, and Irreversible Deformation in Condensed Matter

Vol.689: W. Pötz, J. Fabian, U. Hohenester (Eds.), Quantum Coherence

Vol.690: J. Asch, A. Joye (Eds.), Mathematical Physics of Quantum Mechanics

Vol.691: S. S. Abdullaev, Construction of Mappings for Hamiltonian Systems and Their Applications

Vol.692: J. Frauendiener, D. J. W. Giulini, V. Perlick (Eds.), Analytical and Numerical Approaches to Mathematical Relativity

Vol.693: D. Alloin, R. Johnson, P. Lira (Eds.), Physics of Active Galactic Nuclei at all Scales

Vol.694: H. Schwoerer, J. Magill, B. Beleites (Eds.), Lasers and Nuclei

Vol.695: J. Dereziński, H. Siedentop (Eds.), Large Coulomb Systems

Vol.696: K.-S. Choi, J. E. Kim, Quarks and Leptons From Orbifolded Superstring

Vol.697: E. Beaurepaire, H. Bulou, F. Scheurer, J.-P. Kappler (Eds.), Magnetism: A Synchrotron Radiation Approach

Vol.698: S. Bellucci (Ed.), Supersymmetric Mechanics – Vol. 1

Vol.699: J.-P. Rozelot (Ed.), Solar and Heliospheric Origins of Space Weather Phenomena

Vol.700: J. Al-Khalili, E. Roeckl (Eds.), The Euroschool Lectures on Physics with Exotic Beams, Vol. II

Vol.701: S. Bellucci, S. Ferrara, A. Marrani, Supersymmetric Mechanics – Vol. 2

Vol.702: J. Ehlers, C. Lämmerzahl, Special Relativity